



Outils chimiques pour l'étude de la 5-hydroxyméthylcytosine, une base modifiée de l'ADN

Tom Gerard-Hirne

► To cite this version:

Tom Gerard-Hirne. Outils chimiques pour l'étude de la 5-hydroxyméthylcytosine, une base modifiée de l'ADN. Chimie analytique. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066081 . tel-01163606

HAL Id: tel-01163606

<https://theses.hal.science/tel-01163606>

Submitted on 15 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

ED406

Laboratoire des Biomolécules – UMR 7203 – UPMC/ENS/CNRS

**Outils chimiques pour l'étude de la
5-hydroxyméthylcytosine, une base modifiée de l'ADN**

Par Tom Gérard-Hirne

Thèse de doctorat de Chimie

Dirigée par Dominique Guianvarc'h

Présentée et soutenue publiquement le 9 Février 2015

Devant un jury composé de :

Alain Burger, Professeur, Rapporteur

Mélanie Ethève-Quelquejeu, Professeur, Rapporteur

Casimir Blonski, Directeur de recherche, Examineur

Olivier Lequin, Professeur, Examineur

Carole Saintomé, Maître de conférences, Examineur

Dominique Guianvarc'h, Maître de conférences, Examineur

Remerciements

Je remercie vivement les professeurs Alain Burger et Mélanie Ethève-Quelquejeu d'avoir accepté d'être les rapporteurs de cette thèse. Je remercie aussi le docteur Casimir Blonski et le professeur Olivier Lequin d'évaluer ce travail et d'avoir accepté d'être membres du jury.

Je tiens à remercier Solange Lavielle et Sandrine Sagan de m'avoir accueilli au sein du Laboratoire des Biomolécules pour effectuer mon travail de thèse.

Je remercie Dominique Guianvarc'h pour son encadrement au long de ces années de thèse et son soutien constant. Tu m'as appris énormément et tu m'as guidé pendant ces travaux de thèse, et j'ai pris beaucoup de plaisir à travailler à tes côtés. Merci encore !

J'exprime aussi ma reconnaissance à Carole Saintomé pour m'avoir encadré lors de mes travaux au laboratoire « Régulation et Dynamique du génome », pour m'avoir enseigné tout un pan de techniques que je ne maîtrisais pas et pour ton franc-parler et ta bonne humeur qui m'ont accompagnés pendant mes travaux au museum.

Je souhaite aussi remercier Jean-François Riou, Emmanuelle Delagoutte, Armel Millet et Loic Perron et tous les membres du laboratoire « Régulation et Dynamique du génome » pour leur aide et leur accueil chaleureux lors de ces manipulations au muséum.

La synthèse automatique des oligonucléotides a été réalisée dans l'équipe « Molécules Bioactives » de l'Institut de Chimie de Nice. Je remercie Rachid Benhida de m'avoir accueilli au sein de cette équipe, Benoît Michel pour son aide indispensable à la bonne réalisation de ces synthèses, ainsi que Nicolas et Hella pour les discussions et les bons moments passé à Nice.

Un grand merci à tout ceux avec qui j'ai pu partager mes années de thèse : Claire, Nicolas, Karine, Sophie, Thierry, Françoise, Fabienne, Emmanuelle, Gérard, Gilles, Jesus, Mehdi, Pierre, Astrid, Frédéric, Sébastien, Séverine, Faustine, Thibault, Dimitra, Lucrèce, Chérine,

Jean, Adrian, Lucie, Lucille, Anais H. et Anais T., Ludovic, Bruno et Shahid mais aussi aux autres membres du laboratoire des biomolécules.

Enfin, je souhaite remercier ma famille et mes amis, pour leur soutien tout au long de cette thèse.

Liste des Abréviations

2-OG	2-Oxoglutarate
5azadC	5-Azadésoxycytidine
5caC	5-Carboxycytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxyméthylcytosine
5mC	5-Méthylcytosine
⁵¹ dU	5-Iododésoxyuridine
^{4S} T	4-Thiothymidine
Ac	Acétyl
AdoMet	<i>S</i> -Adénosyl-L-Méthionine
ADN	Acide désoxyribonucléique
AIBN	Azobisisobutyronitrile
ARN	Acide ribonucléique
BER	Réparation par excision de base
BGT	β-Glucosyltransférase du phage T4
BS	Séquençage au bisulfite
CCM	Chromatographie sur couche mince
CHO	Cellules d'ovaires de hamster de chine
CpG	Dinucléotide CG
DBCO	Dibenzylcyclooctine
DBU	1,8-Diazabicyclo(5.4.0)undéc-7-ène
DIEA	<i>N,N</i> -Diisopropyléthylamine
DMAP	4-Diméthylaminopyridine
DMF	<i>N,N</i> -Diméthylformamide
DMSO	Diméthylsulfoxyde
DMT	Diméthoxytrityl
DNMT	Méthyltransférase de l'ADN
DSBH	Double Stranded Beta Helix
ESI	Ionisation par électronébulisation
HATU	1-[Bis(diméthylamino)méthylène]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate

HEPES	Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HPLC	Chromatographie en phase liquide à haute performance
JBP	J-Binding Protein
LC- MS/MS	Chromatographie en phase liquide couplée à la spectrométrie de masse en tandem
LED	Diode électroluminescente
MALDI-TOF	Matrix assisted laser desorption ionisation – Time of Flight
MBD	Methyl Binding Domain
Me	Méthyl
MRM	Multiple Reaction Monitoring
NBS	N-Bromosuccinimide
ODN	Oligodéoxynucléotide
PAL	Photoaffinity labeling, Photomarquage d'affinité
PEG	Polyéthylène glycol
RPA	Protéine de réplication A
SDS	Dodécyl sulfate sodium
SDS – PAGE	Electrophorèse dénaturante en gel de polyacrylamide en présence de SDS
SSB	Protéine de liaison à l'ADN simple brin
TBAF	Fluorure de tétra- <i>n</i> -butylammonium
TBE	Tris, Borate, EDTA
TBDMS	<i>Tert</i> -butyldiméthylsilyle
TDG	Thymine DNA Glycosylase
TET	Ten eleven Translocation
TFA	Acide trifluoroacétique
T _m	Température de fusion
TRIS	2-amino-2-hydroxyméthyl-1,3-propanediol
UDP-Glucose	Uridine diphosphate glucose
UV	Ultraviolet

Table des matières

Remerciements	1
Liste des Abréviations.....	3
Table des matières	5
Introduction Générale.....	11
1 La méthylation de l'ADN : une des composantes de la régulation épigénétique	13
2 La 5-hydroxyméthylcytosine : de l'ombre à la lumière	16
2.1 L'hydroxyméthylation des bases nucléiques dans l'ADN : l'état des lieux avant 2009	17
2.1.1 Le cas du bactériophage T4.....	17
2.1.2 La 5hmC chez les mammifères	19
2.1.3 La base J des Trypanosomatidae	20
2.2 La redécouverte de la 5-hydroxyméthylcytosine	22
3 Formation de la 5-hydroxyméthylcytosine par les enzymes TET.....	23
3.1 La famille des protéines TET	23
3.2 Cycle catalytique des enzymes TET	24
3.3 Substrats des enzymes TET.....	24
3.4 L'interaction ADN/TET : données structurales et biochimiques.....	25
4 Stabilité et mutagenicité de la 5hmC, 5fC et 5caC	28
5 Distribution des dérivés oxydés de la 5mC chez le mammifère dans l'espace et dans le temps	30
5.1 Les niveaux de 5hmC varient en fonction de la nature des tissus.....	30
5.2 La 5fC et la 5caC sont présentes en très faible quantité dans l'ADN	31
5.3 Cartographie des dérivés oxydés de la 5mC dans le génome	32
5.4 Variation des niveaux globaux de 5hmC dans le temps.....	33
6. Les enzymes TET et la 5hmC chez d'autres organismes.....	36
7. Rôles biologiques de l'oxydation des 5mC par les protéines TET	37
7.1 La 5hmC comme intermédiaire dans un processus de déméthylation active de la 5mC	37
7.1.1 Interférence dans la maintenance de la méthylation de l'ADN	38

7.1.2 Déméthylation par les voies de réparation de l'ADN	39
7.1.3 Déméthylation enzymatique « directe ».....	41
7.2 La 5hmC et ses dérivés oxydés : des marqueurs épigénétiques ?	46
7.2.1 Identification de protéines interagissant avec les différentes modifications de la cytosine.....	47
7.2.2. Etudes de quelques protéines de liaison à l'ADN méthylé	47
7.2.4 Recherche de protéines interagissant spécifiquement avec 5hmC et ses dérivés.	49
8. L'implication de la 5hmC dans des processus biologiques et des pathologies : les dernières avancées.....	53
8.1 La 5hmC dans le développement	53
8.1.1 Variation des taux de 5hmC dans le développement embryonnaire	53
8.1.3 La 5hmC et les protéines TET dans la reprogrammation de cellules souches pluripotentes.....	55
8.2. Mécanismes de régulation des protéines TET.....	56
8.3 La 5hmC et les protéines TET dans le cancer	56
8.4 La 5hmC dans les maladies neurodéveloppementales et neurodégénératives	58
9. Conclusion.....	60
10. Projet de thèse	62
Quantification globale par spectrométrie de masse : de la méthylation à l'hydroxyméthylation.....	65
1 Quantification de la méthylation de l'ADN par spectrométrie de masse en tandem avec ionisation electrospray couplée à une chromatographie liquide	68
2 Validation d'une technique de quantification de la méthylation globale d'échantillons d'ADN.....	71
3 La spectrométrie de masse pour l'étude de la 5hmC	73
3.1 Les études déjà réalisées	73
3.2 Mise au point de la technique au laboratoire.....	74
Outils chimiques pour le marquage de la 5-hydroxyméthylecytosine dans une séquence d'ADN.....	80
1 Etat des lieux en 2014 des techniques développées pour l'étude de séquences d'ADN contenant la 5hmC.....	83

1.1. Reconnaissance des séquences 5hmC par des anticorps	83
1.2 Enrichissement de séquences par marquage chimio-enzymatique des 5hmC	84
1.2.1 Hydroxyméthylcytosine selective affinity labeling (hMe-Seal)	85
1.2.2 Interaction avec la protéine JBP1	86
1.2.3 La méthode GLIB (Glucosylation, periodate oxydation, biotinylation)	87
1.3 Enrichissement de séquences par marquage chimique des 5hmC	88
1.4 Cartographie par les enzymes de restriction	89
1.5 Les techniques de séquençage	92
1.5.1 Rappels sur le séquençage au bisulfite	92
1.5.2 Les techniques dérivées du séquençage au bisulfite	94
1.5.2.1 Les méthodes de séquençage OxBS et RedBS	94
1.5.2.2 La méthode de séquençage Tet-Assisted Bisulfite sequencing (TAB-seq) ..	95
1.5.2.3 Les méthodes de séquençage fCAB-seq et CAB-seq	96
1.5.3 Le séquençage par nanopores	99
1.5.4 Le séquençage en temps réel de molécule unique (Single Molecule Real-Time sequencing, SMRT)	100
1.6 Les méthodes de quantification globale par marquage radioactif ou fluorescent ...	102
2 Développement de méthodes de marquage de la 5-hydroxyméthylcytosine	104
2.1 Marquage enzymatique des 5hmC	104
2.1.1 Synthèse de la séquence oligonucléotidique	107
2.1.2 Synthèse de la molécule UDP-6-azido-glucose	114
2.1.3 Mise au point de la réaction de chimie « click »	117
2.1.4 Obtention de la BGT	119
2.1.5 Tentatives de transfert du glucose modifié sur le substrat modèle	122
2.1.6 Conclusion	124
2.2 Marquage chimique des 5hmC	125
2.2.1 Mise au point des conditions d'oxydation et de fonctionnalisation de la 5hmC	126
2.2.1.1 Synthèse de la 5-hydroxyméthyl-2'-désoxycytidine	126

2.2.1.2 Mise au point des conditions d'oxydation de l'alcool allylique de la 5hmdC	129
2.2.1.3 Mise au point des conditions de fonctionnalisation de la 5hmdC par la biotine-hydrazine.....	131
2.2.2 Oxydation et fonctionnalisation d'un oligonucléotide modèle	134
2.2.3 Conclusion.....	137
Sondes photoactivables pour l'étude de protéines interagissant avec les bases rares de l'ADN.....	140
1 Rappels bibliographiques	143
1.1 Le photomarquage d'affinité	143
1.2 Les principaux groupements photoactivables	145
1.2.1 Les azotures d'aryle	148
1.2.2 Benzophénone	150
1.2.3 Diazirine	152
1.3 Analogues de nucléosides	155
1.3.1 Analogues thiolés d'acides nucléiques.....	155
1.3.2 Dérivés halogénés de l'uridine	158
1.4 Introduction des groupements photoactivables dans des sondes oligonucléotidiques	159
1.4.1 Synthèse enzymatique de sondes oligonucléotidiques photoactivables.....	160
1.4.2 Synthèse chimique de sondes oligonucléotidiques photoactivables	161
1.4.3 Stratégie pré-synthétique : introduction du groupement photoactivable via les synthons phosphoramidites	161
1.4.3.1 Introduction directe de nucléobases photoactivables	162
1.4.3.2 Incorporation d'agent de photomarquage via des bras espaceurs sur le nucléoside.....	162
1.4.4 Introduction du groupement photoactivable par la fonctionnalisation post-synthétique	166
1.4.4.1 Méthode du nucléoside convertible.....	167
1.4.4.2 Réactions bioorthogonales	171

2 Conception, synthèse et évaluation de sondes photoactivables pour l'identification de protéines partenaires de la 5hmC	174
2.1 Conception des sondes	174
2.2 Synthèse des sondes photoactivables	176
2.2.1 Synthèse des amines comportant les groupements photoactivables	178
2.2.2 Synthèse des oligonucléotides en phase solide	179
2.2.3. Fonctionnalisation post-synthétique des oligonucléotides	180
2.2.4. Récapitulatif des oligonucléotides synthétisés	182
2.3 Etude de stabilité des duplexes étudiés	185
2.3.1 Expériences de gel retard	185
2.3.2 Expériences de température de fusion	187
2.4 Mise au point des conditions de photomarquage avec une protéine modèle	189
2.4.1 Etude des complexes duplex-DNMT3a	189
2.4.2 Etude des complexes simple brin-RPA	192
2.4.3 Essais d'irradiation avec des LEDs	199
2.5 Essais de photomarquage en présence d'un mélange de protéines	203
2.6 Mise au point des conditions de capture sur bille de streptavidine	207
2.7 Essais de photomarquage sur des extraits nucléaires	211
3 Conclusion et perspectives	214
Conclusion générale	217
Partie expérimentale	221
Bibliographie.....	249
Table des illustrations	277
Table des tableaux	285

Chapitre I

Introduction Générale

En 2009, un nouvel acteur de la scène de l'épigénétique entre en jeu : il s'agit de la 5-hydroxyméthylcytosine, qualifiée très rapidement de sixième base de l'ADN. Comment cette base a-t-elle été caractérisée ? Pourquoi a-t-elle échappé jusque-là à l'attention de la communauté scientifique ? Comment se forme-t-elle et quel est son rôle biologique ? Au cours de cette introduction bibliographique, nous allons aborder ces différentes questions avant de présenter dans les chapitres suivants notre contribution à l'étude de cette « nouvelle » base de l'ADN.

1 La méthylation de l'ADN : une des composantes de la régulation épigénétique

L'histoire de l'épigénétique est liée à celle de l'étude de l'évolution et du développement et part d'un constat simple : si toutes les cellules d'un organisme possèdent la même information génétique, il existe des différences remarquables entre les différents types de cellules, résultant de différences dans l'expression des gènes. Dans les années 40, Waddington propose le mot d'épigénétique pour définir ces processus de régulation de l'information génétique, d'après le préfixe grec épi, signifiant « au-dessus »¹. Depuis, la compréhension de l'épigénétique a évolué et la définition de l'épigénétique englobe les changements dans l'expression des gènes ou des phénotypes cellulaires qui ne sont pas la conséquence de changements dans la séquence d'ADN. Ces changements sont réversibles et, dans certains cas, héréditaires.

De nombreux phénomènes épigénétiques ont été mis à jour, comme les modifications des histones, des protéines qui vont réguler l'état de la chromatine et son accessibilité, ou bien la méthylation de l'ADN. Chez les mammifères, la méthylation de l'ADN a lieu quasi-exclusivement à la position 5 du cycle pyrimidique de cytosines suivies d'une guanine dans la séquence d'ADN, appelés dinucléotides CpG. Cette réaction de méthylation est catalysée par les enzymes de la famille des ADN méthyltransférases (DNMT), à l'aide du cofacteur S-Adénosyl-L-Méthionine (AdoMet), qui sert de donneur de méthyle (Figure 1).

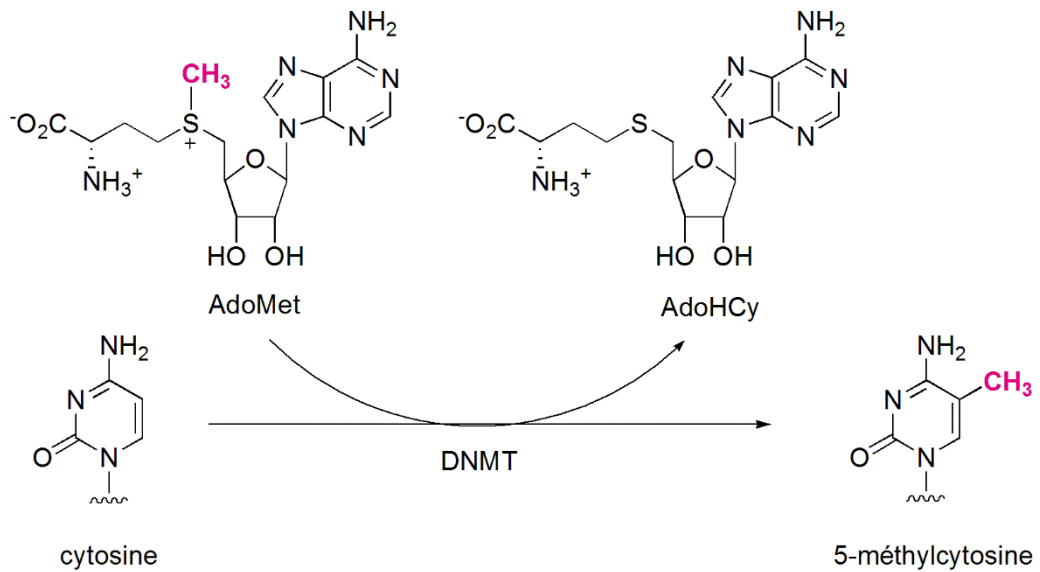


Figure 1. La réaction de méthylation de l'ADN par les DNMT.

Les enzymes DNMT utilisent un mécanisme commun pour effectuer la réaction de méthylation (Figure 2). La cytosine cible bascule depuis la double hélice pour être insérée dans la poche catalytique de l'enzyme, puis une cystéine catalytique s'additionne sur la position 6 du cycle pyrimidique de la cytosine activant ainsi la position 5. Le groupement méthyle présent sur le cofacteur AdoMet pourra ainsi ensuite être transféré sur la position 5 de la cytosine, puis une réaction d'élimination finale va permettre de libérer la cystéine catalytique et d'accéder à la cytosine méthylée².

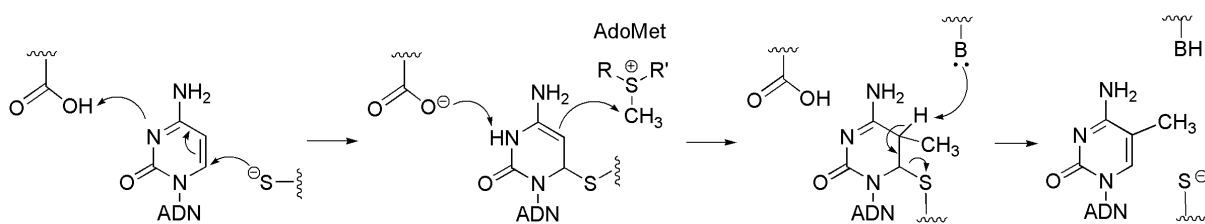


Figure 2. Mécanisme de la méthylation des cytosines catalysé par les DNMTs.

Les DNMT humaines peuvent être classées en deux groupes différents selon la fonction de méthylation qu'elles vont assurer. La méthylation *de novo* est ainsi assurée par les DNMT3, tandis que la maintenance de la méthylation préalablement établie est assurée par la DNMT1.

Le groupe DNMT3 comprend deux enzymes actives, DNMT3A et DNMT3B, ainsi qu'une troisième protéine catalytiquement inactive, DNMT3L (pour DNMT3 like). DNMT3L peut s'associer avec DNMT3A et 3B pour stabiliser la conformation de leur site actif et augmenter leur activité. C'est ce groupe d'enzymes qui va permettre d'établir les profils de méthylation initiaux, « de novo » (Figure 3).

La DNMT1 permet d'assurer la maintenance de la méthylation au cours des divisions cellulaires. En effet, lors de la réplication d'un ADN méthylé sur les deux brins, la synthèse d'un nouveau brin d'ADN entraîne la formation de sites CpG méthylés sur un seul brin (ADN hémiméthylé). La DNMT1 possède une affinité plus grande pour les CpG hémiméthylés que pour les CpG non modifiés, et permet donc de restaurer les profils de méthylation de l'ADN sur le nouveau brin formé lors de la division cellulaire (Figure 3).

Figure 3. La méthylation de novo par les DNMT3 et la méthylation de maintenance par DNMT1 dans l'ADN³.

Dans la famille des DNMTs, il existe également une enzyme nommée DNMT2 qui n'est en fait pas capable de catalyser la méthylation des cytosines dans l'ADN. Son rôle biologique a longtemps fait l'objet de spéculation, jusqu'à ce que l'on mette en évidence le fait qu'elle peut méthyler la cytosine présente en position 38 de l'ARN de transfert de l'acide aspartique (tRNA^{Asp})⁴ par un mécanisme similaire à celui des ADN méthyltransferase⁵. C'est, à ce jour, la seule fonction trouvée pour la DNMT2.

Les dinucléotides CpG ne sont pas uniformément répartis dans le génome humain et sont principalement regroupés dans des séquences comportant de nombreux CpG, appelées îlots CpG, présents dans les régions promotrices des gènes, dans les transposons et dans les séquences répétées des chromosomes^{6,7}. La présence des 5mC dans les régions promotrices a un rôle dans la transcription des gènes associés. En effet, le groupement méthyle présent sur la cytosine empêche d'une part la liaison des facteurs de transcription sur l'ADN et d'autre part permet de recruter des partenaires protéiques spécifiques des 5mC qui vont ensuite permettre la modification des histones et induire le compactage de la chromatine. La 5mC est donc considérée comme une modification épigénétique répressive, qui empêche l'expression de certains gènes (Figure 4).

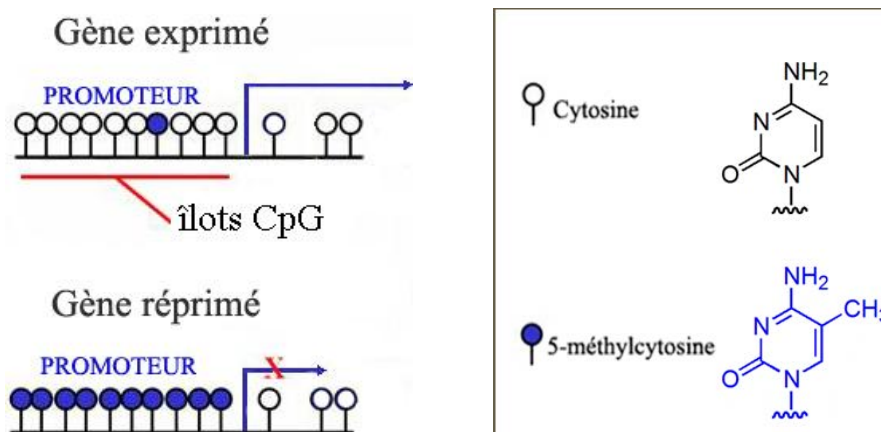


Figure 4. La régulation de l'expression des gènes est étroitement liée à la méthylation de l'ADN.

La méthylation de l'ADN intervient ainsi dans les processus de régulation de l'expression des gènes, mais également dans différents processus biologiques comme la différenciation cellulaire, l'inactivation d'un des deux chromosomes X chez les femelles des mammifères, l'empreinte parentale, qui consiste en la répression d'une copie d'un gène selon qu'elle est d'origine paternelle ou maternelle ou encore dans la stabilité chromosomique⁸. Des variations dans les profils de méthylation de l'ADN sont également impliquées dans certaines pathologies parmi lesquelles certains cancers. En effet, dans les cancers, une hypométhylation globale est souvent associée, de manière contre-intuitive, à une augmentation aberrante de la méthylation dans certaines parties de l'ADN. Ainsi, on retrouve souvent une hyperméthylation des promoteurs de gènes suppresseurs de tumeurs qui va empêcher leur expression. L'un des premiers exemples d'une telle répression a été détecté dans le rétinoblastome⁹. La méthylation de l'ADN est ainsi un processus dynamique et réversible, et constitue un mécanisme épigénétique à part entière.

2 La 5-hydroxyméthylcytosine : de l'ombre à la lumière

La 5mC est donc une base fréquemment rencontrée dans l'ADN et son rôle biologique primordial lui a valu d'être qualifiée de 5^{ème} base de l'ADN. Toutes les autres bases modifiées étaient considérées jusque récemment comme rares et anormales, se formant par exemple suite à un dommage oxydatif touchant accidentellement un nucléotide. La 5-hydroxyméthylcytosine (5hmC) entrait dans cette catégorie jusqu'à ce que deux publications parues en 2009 remettent en question cette idée^{10,11}. Elles montrent en effet que, d'une part, la 5hmC est abondante dans certains types cellulaires et que d'autre part, la

formation de 5hmC est un processus actif dirigé par des enzymes spécialisées (TET1-3) qui catalysent l'oxydation de la 5mC, pour former la 5hmC (Figure 5).

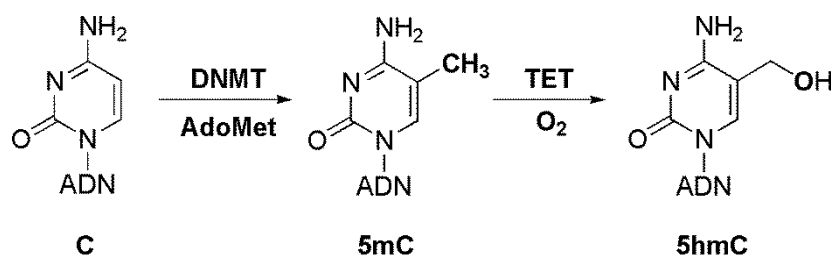


Figure 5. Biosynthèse de la 5-hydroxyméthylcytosine, la 6^{ème} base de l'ADN.

Depuis la parution de ces travaux, la 5-hydroxyméthylcytosine est devenue une composante incontournable des processus épigénétiques. Nous allons présenter dans un premier temps ce que l'on connaissait de cette base avant 2009 et de ses analogues considérés comme marginaux ou exotiques avant ce que l'on peut considérer comme sa « redécouverte » que nous détaillerons également.

2.1 L'hydroxyméthylation des bases nucléiques dans l'ADN : l'état des lieux avant 2009

Si la découverte majeure de 2009 a remis la 5hmC sur le devant de la scène, sa présence ainsi que celle de nucléobases pyrimidiques modifiées similaires avait déjà été observée dans différents organismes et suite à divers processus soit chimiques soit enzymatiques que nous allons présenter.

2.1.1 Le cas du bactériophage T4

Afin de protéger leur génome de l'action des enzymes de restriction des bactéries qui sont leurs cibles, certains virus ont développé des stratégies qui passent notamment par la modification de bases de leur ADN. Chez le phage T4, la totalité des cytosines est ainsi remplacée par la 5-hydroxyméthylcytosine (5hmC), et plus tard modifiée par des α - et β -glucosyltransférases (AGT ou BGT) en présence d'uridine diphosphate glucose (UDP-Glucose)¹²⁻¹⁴ (Figure 6). L'incorporation des 5hmC dans l'ADN se fait pendant la répllication de celui-ci. En effet, la dCMP hydroxyméthylase du phage T4 catalyse la formation de la 2'-

désoxy-5-hydroxyméthylcytidine-5'-monophosphate (d5hmCMP) à partir de la 2'-désoxycytidine-5'-monophosphate (dCMP) (Figure 6). La 5hmdCMP est alors phosphorylée en d5hmCTP qui est ensuite utilisée comme monomère constitutif lors de la réplication de l'ADN. Du point de vue mécanistique, la réaction est similaire à la méthylation des cytosines dans l'ADN en position 5 *via* l'activation par une cystéine en position 6 du cycle pyrimidine, à la différence que le méthylène transféré ne provient pas de l'AdoMet mais du méthylènetétrahydrofolate (THF) (Figure 6).

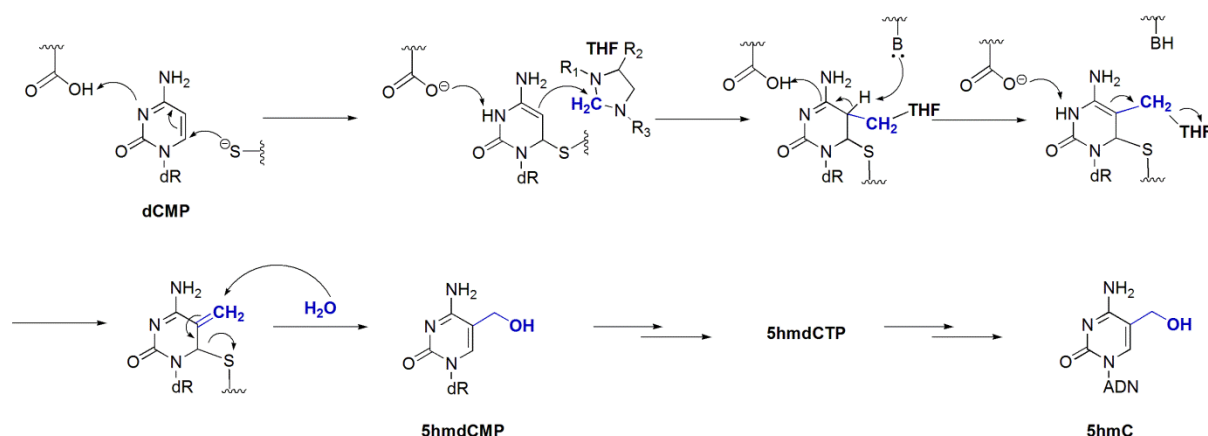


Figure 6. La réaction catalysée par la dCMP hydroxyméthylase permettant la conversion de dCMP en d5hmCMP chez le phage T4.

Les 5hmC présentes dans l'ADN du phage T4 sont ensuite modifiées par le transfert d'un groupement glucose sur l'alcool allylique en position 5 (Figure 7). Cette réaction est catalysée par des glucosyltransferases (AGT ou BGT) et permet de protéger complètement l'ADN du phage vis-à-vis des enzymes de restriction de l'organisme infecté.

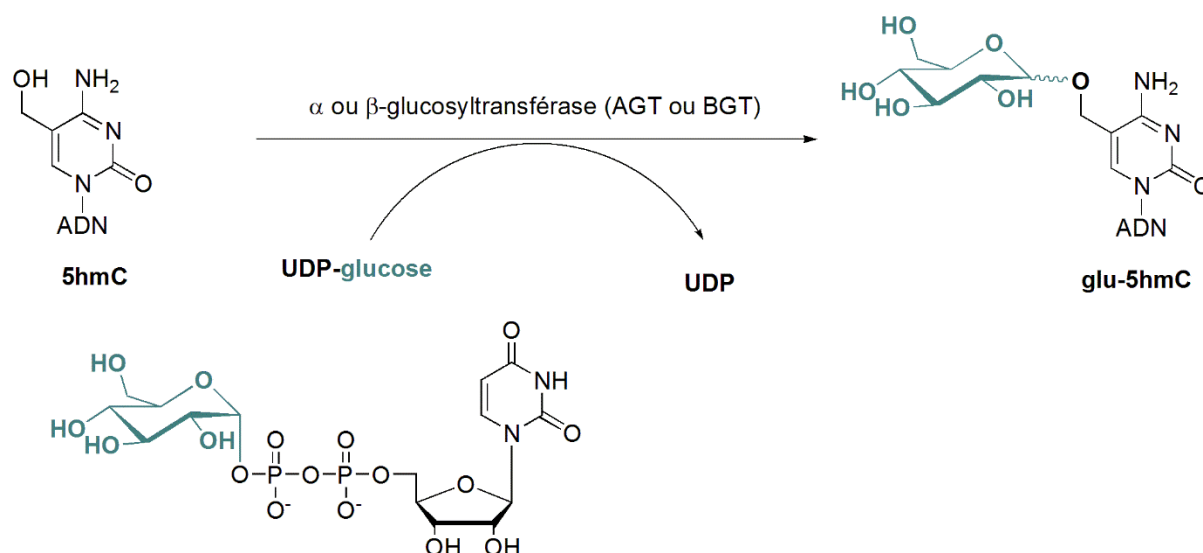


Figure 7. Glucosylation des 5-hydroxyméthylcytosines par les enzymes AGT ou BGT chez le phage T4.

2.1.2 La 5hmC chez les mammifères

Chez les mammifères, la 5hmC a été identifiée pour la première fois dans des extraits de cerveaux et de foie de rats dans les années 1970, à des niveaux très élevés (15% et 17%, respectivement, par rapport aux cytosines totales)¹⁵. Les niveaux de 5hmC rapportés ainsi que le fait que la même publication n’observait pas de présence de 5mC laissent néanmoins penser que le traitement de l’ADN par une solution aqueuse à 88% d’acide formique à 175°C avant analyse a pu dégrader l’ADN conduisant à des taux de 5hmC erronés.

La première observation de la 5hmC dans l’ADN de mammifères a ainsi été ignorée pendant près de 40 ans, la communauté scientifique considérant que la présence de faibles quantités de 5hmC était due aux conséquences des dégâts oxydatifs sur les 5-méthylcytosines présentes dans l’ADN. Comme on peut le voir dans la figure 8, l’oxydation de la 5mC présente dans l’ADN en présence d’espèces réactives oxygénées entraîne la formation du radical peroxy ainsi que de l’hydroperoxyde qui se transforment ultimement en différentes espèces oxydées des cytosines, comme par exemple la 5hmC mais également la 5-formylcytosine (5fC)^{16–18}.

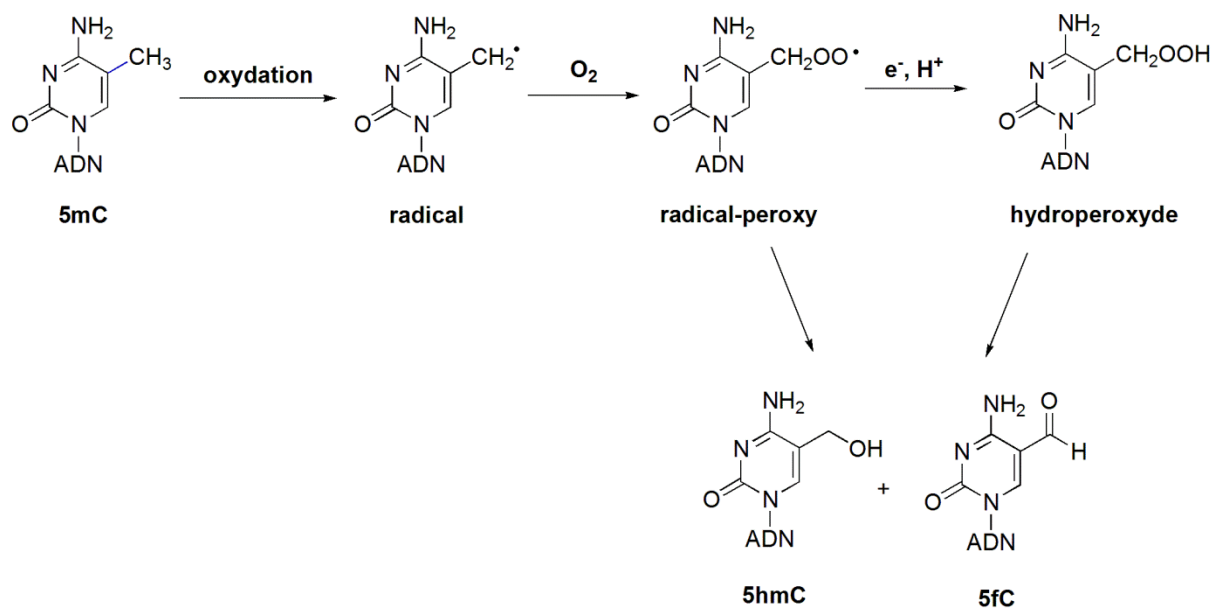
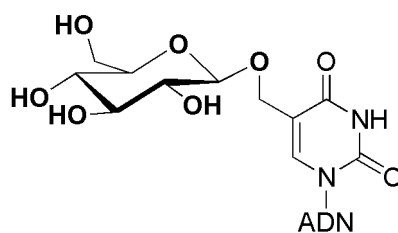


Figure 8. Formation de dérivés modifiés de la 5mC consécutivement à des dégâts oxydatifs.

2.1.3 La base J des Trypanosomatidae

La famille des Trypanosomatidae est composée de parasites eucaryotes des animaux, dont la transmission à l'homme entraîne l'apparition de maladies graves. Ainsi, *Trypanosoma Brucei* est responsable de la maladie du sommeil et *Trypanosoma Cruzi* de la maladie de Chagas. La famille des *Leishmania* est quant à elle responsable de la transmission des leishmanioses. Ces parasites possèdent une modification rare de l'ADN, la β-D-glucopyranosyloxymethyluracile, ou base J (Figure 9), une structure très similaire à celle trouvée chez le phage T4. La base J représente 1% des thymines présentes dans l'ADN, et est principalement présente dans les séquences répétées ainsi que dans les sites d'initiation et de terminaison de la transcription par l'ARN polymérase II (Pol II). La base J est considérée comme une modification épigénétique chez ces parasites et a donc une voie de biosynthèse et un rôle bien différent de la base glu-5hmC du phage T4.



Base J

Figure 9. Structure de la base J retrouvée exclusivement dans l'ADN des Trypanosomatidae.

Ainsi, contrairement à ce qui se passe lors de la biosynthèse de la 5hmC chez le phage T4, la formation de la base J, qui passe par un intermédiaire 5-hydroxyméthyl-uridine (5hmU) similaire à la 5hmC, a lieu sur des brins d'ADN déjà synthétisés. Cette modification de l'ADN se fait en deux étapes enzymatiques au niveau de certaines thymidines. Dans un premier temps, les enzymes JBP1 et JBP2 (J Binding Protein), appartenant à la famille des dioxygénases à Fe (II) dépendante du 2-oxoglutarate (2-OG), vont catalyser l'hydroxylation du groupement méthyle présent sur la position 5 de la thymidine conduisant à la formation de 5hmU (Figure 10)¹⁹. Bien que n'ayant été démontré que récemment, le rôle des enzymes JBP1 et JBP2 dans l'oxydation de la thymine était soupçonné depuis longtemps. Dans un deuxième temps, la 5hmU est glycosylée et c'est tout récemment que cette étape a été élucidée suite à la caractérisation de la glucosyltransférase, nommée JGT et du donneur de glucose, l'UDP-glucose. Selon un processus similaire à ce qui se passe dans la phage T4, la JGT catalyse le transfert d'un glucose sur l'alcool allylique préalablement formé conduisant à la formation de la base J (Figure 10)²⁰.

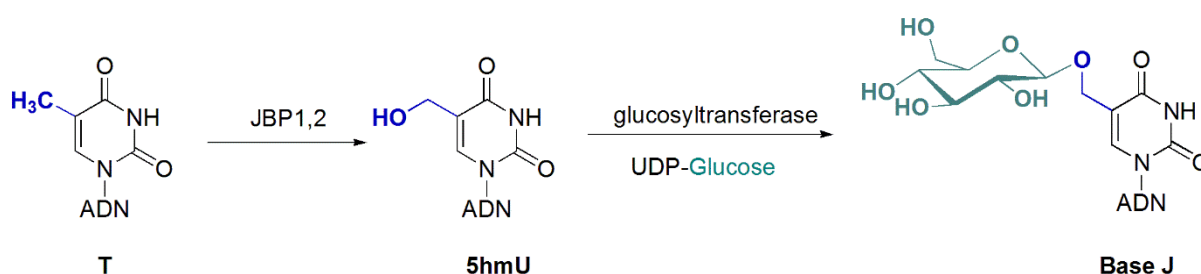


Figure 10. Formation de la base J dans l'ADN des Trypanosomatidae.

2.2 La redécouverte de la 5-hydroxyméthylcytosine

La mise en évidence d'enzymes capables de catalyser l'oxydation du groupement méthyle de la thymine, qui a une structure proche de la cytosine méthylée, et l'implication de ces nouvelles modifications dans des processus épigénétiques ont récemment éveillé l'intérêt de la communauté scientifique. La recherche d'homologues des JBP à partir de la structure de leur domaine catalytique supposé va permettre de mettre à jour différentes familles de dioxygénases à Fe (II) dépendantes du 2-oxoglutarate et capables de catalyser des réactions d'oxydations sur des bases nucléiques²¹. Parmi celles-ci sont notamment identifiées les enzymes Ten Eleven Translocation (TET), très proches structuralement des enzymes JBP.

En 2009, deux publications sorties simultanément dans *Science* ont ensuite permis la redécouverte de la présence de la 5hmC chez le mammifère, et l'implication des protéines TET dans sa formation. En effet, dans la lancée de la recherche d'homologues des JBP, le groupe d'Anjana Rao montre que la protéine TET1 est capable de catalyser la formation de 5hmC à partir de 5mC présentes dans l'ADN (Figure 5). De plus, l'utilisation d'une combinaison de digestion de l'ADN et de séparation de nucléotides radiomarqués par chromatographie sur couche mince leur permet de détecter la 5hmC dans des cellules souches embryonnaires de souris¹⁰. Skirmantas Kriaucionis et Nathaniel Heintz rapportent, quant à eux, la détection et l'identification de 5hmC dans les cellules de Purkinje et des cellules de cerveau de souris par des techniques utilisant la chromatographie sur couche mince et la spectrométrie de masse¹¹.

Depuis 2009 et la découverte d'enzymes catalysant sa formation, la 5-hydroxyméthylcytosine est sur le devant de la scène. De nombreuses études se sont consacrées à approfondir les connaissances concernant sa formation, ainsi qu'à élucider son rôle biologique. Les paragraphes suivants consisteront en une synthèse de la littérature sur la formation de la 5hmC par les enzymes TET, sa présence dans les différents types de tissus et dans les séquences d'ADN, ainsi que les différentes hypothèses sur son rôle biologique et son implication dans certaines pathologies. La redécouverte de cette base a aussi impulsé le développement de nombreuses méthodes pour la détecter, la quantifier et la séquencer, qui seront présentées dans l'introduction du chapitre III.

3 Formation de la 5-hydroxyméthylcytosine par les enzymes TET

3.1 La famille des protéines TET

Chez les mammifères, la famille des enzymes TET comprends trois membres, nommés TET1, TET2 et TET3 (Figure 11A). Ces trois membres sont issus de la triplication d'un gène ancestral des TET. Après cette triplication, le gène codant pour TET2 a subi une inversion chromosomique qui a entraîné la séparation de son domaine CXXC, désormais encodé par un gène spécifique nommé IDAX (ou CXXC4), du reste de l'enzyme (Figure 11B).

Figure 11. La famille des protéines TET chez les mammifères. A) Séquence schématique des enzymes TETs humaines. B) Evolution des enzymes TET chez les mammifères²².

On distingue deux domaines importants :

- Les domaines CXXC *N*-terminaux sont des domaines à doigts de Zn^{2+} qui vont contenir une ou plusieurs séquences caractéristiques de type $CGXCXXC(X)_nC$, où C est une cystéine, G une glycine et X peut être n'importe quel acide aminé. Les domaines CXXC sont généralement présents dans les protéines associées à la réorganisation de la chromatine, comme les methyl-binding protein (MBD), ou bien DNMT1. Chez les enzymes TET, les domaines CXXC sont de petite taille, entre 56 et 58 acides aminés, et comprennent deux séquences caractéristiques.

- Un domaine catalytique comprenant un domaine riche en cystéines suivi d'un domaine *C*-terminal de type double-brin hélice β (double-stranded beta helix ou DSBH). Le domaine riche en cystéines est conservé dans la famille des protéines TET et le domaine DSBH est le cœur catalytique classique des dioxygénases à Fe (II) et 2-OG. Il comprend un motif $HX...H$, où X est soit un acide aspartique soit un acide glutamique, responsable de la complexation de l'atome de fer impliqué dans le processus catalytique. De plus, une insertion de faible complexité, dont la séquence est peu conservée à travers les différentes protéines

TET, est présente au milieu de ce domaine. Cette insertion sépare les deux moitiés du motif de complexation du fer HX...H.

3.2 Cycle catalytique des enzymes TET

Le mécanisme supposé de l'oxydation de la 5mC en 5hmC, caractéristique des dioxygénases à Fe (II) et dépendantes du 2-oxoglutarate, est résumé dans la figure 12. La coordination de l'atome de fer par le 2-oxoglutarate et le dioxygène permet d'oxyder le Fe(II) en Fe(IV), en produisant du dioxyde de carbone. C'est ce Fe(IV) qui va ensuite oxyder le substrat (ici, le groupement méthyle présent sur la position 5 de la 5-méthylcytosine) conduisant à l'hydroxylation de celui-ci. Après élimination du succinate et du dioxyde de carbone formé lors de la décarboxylation du 2-oxoglutarate et la coordination d'une nouvelle molécule de 2-oxoglutarate, le cycle catalytique est complété.

Figure 12. Cycle catalytique caractéristique des dioxygénases à Fe(II) dépendantes du 2-oxoglutarate²³.

Après la première observation par le groupe d'Anjana Rao que TET1 était capable de catalyser l'hydroxylation de la 5mC en 5hmC, il a été montré que les trois enzymes de la famille TET possédaient cette même activité catalytique²⁴.

3.3 Substrats des enzymes TET

La 5mC n'est pas le seul substrat des enzymes TETs. En effet, elles sont capables de catalyser l'oxydation successive des dérivés de la 5-méthylcytosine, et de former la 5-formylcytosine (5fC) à partir de la 5hmC, puis de la 5-carboxycytosine (5caC) à partir de la 5fC (Figure 13), sans que l'on sache pour le moment si ces oxydations successives se font par un mécanisme processif ou distributif^{25,26}.

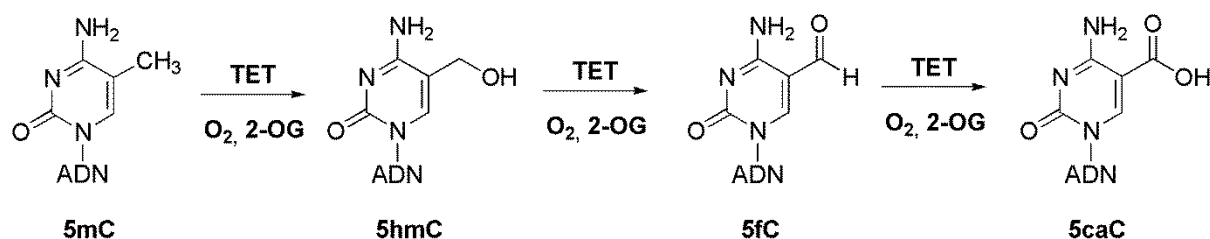


Figure 13. Oxydations successives de la 5mC par les enzymes TET.

Par ailleurs, dans des travaux récents, il a été montré que le domaine catalytique de TET1 est capable de catalyser les oxydations itératives de la 5mC non seulement dans le cas de l'ADN double brin, mais aussi sur de l'ADN simple brin ainsi que sur de l'ARN²⁷. L'activité est la plus forte vis-à-vis de l'ADN double brin, puis de l'ADN simple brin, et enfin de l'ARN. La présence de 5hmC dans l'ARN a ainsi été détectée dans différents tissus et notamment le cerveau humain. Cette capacité des enzymes TET à oxyder l'ARN peut permettre d'expliquer la présence d'un homologue des TET dans le génome de la drosophile, qui ne comporte qu'un homologue de DNMT2 mais pas d'enzymes capables de méthyliser les cytosines de l'ADN²⁸. *In vitro*, les enzymes TET sont aussi capables de catalyser l'hydroxylation de la thymidine en 5hmU, de façon moins efficace que l'oxydation de la 5mC en 5hmC²⁹.

3.4 L'interaction ADN/TET : données structurales et biochimiques

Deux études cristallographiques d'enzymes TET ont actuellement été publiées, et ont apporté des informations précieuses sur la structure de ces enzymes et le rôle des différents domaines.

La première étude décrit la structure de TET2 en complexe avec de l'ADN double brin comprenant un îlot CpG totalement méthylé et la *N*-Oxalylglycine (N-OG), un analogue non réactif du 2-OG en la combinant à l'étude de l'activité catalytique de différents mutants³⁰. Dans la structure obtenue, le domaine DBSH forme un cœur catalytique compact, autour duquel le domaine riche en cystéines va s'enrouler. La complexation d'atomes de zinc par des résidus du domaine riche en cystéines et du domaine DBSH permet de structurer la protéine et

est nécessaire à son activité catalytique. TET2 interagit avec l'ADN par une surface comprenant des résidus basiques et hydrophobes, et cette interaction est stabilisée par des résidus du domaine riche en cystéine. Le domaine DBSH va reconnaître spécifiquement un dinucléotide CpG et va faire basculer une 5mC pour l'insérer dans sa poche catalytique. Comme TET2 ne semble reconnaître que les dinucléotides CpG, cela indique que les protéines TET, si elles sont dirigées vers des régions spécifiques de l'ADN, le sont par leur motif CXXC ou par l'interaction avec d'autres partenaires. A l'intérieur de la poche catalytique de TET2, le groupement méthyle de la 5mC est orienté vers le 2-OG, stabilisé par des interactions avec ses groupements carboxylates, ainsi que vers l'atome de fer, qui est bien complexé par une triade His/Glu/His. Aucun résidu de la protéine n'interagit avec le groupement méthyle de la 5mC. Les auteurs ont confirmé ce résultat dans la même étude en montrant que TET2 avait la même affinité pour un CpG méthylé ou non modifié. Ceci est en accord avec la possibilité pour les enzymes TET d'oxyder successivement la 5mC, la 5hmC et la 5fC et donc la nécessité d'avoir une poche enzymatique capable d'accueillir ces résidus différents. La figure 14 présente une vue rapprochée du site catalytique de TET2 dans deux représentations différentes.

Figure 14. Poche catalytique de l'enzyme TET2, en complexe avec un duplex de l'ADN et N-OG, un analogue du 2-OG ; A) Représentation en bâtons, la distance entre le groupement méthyle et l'atome d'oxygène de la molécule de NOG est indiquée en angström et ne correspond pas à une liaison hydrogène ; B) Représentation en surface.

La seconde étude décrit la structure de NgTet1, un homologue des protéines TET provenant de *Naegleria gruberi*, une amibe flagellée, en complexe avec un substrat totalement méthylé et a permis de confirmer certaines des informations obtenues avec la structure de TET2³¹. Cette enzyme possède une préférence pour les dinucléotides CpG, ainsi que dans une moindre mesure pour les CpA. De manière similaire à TET2, la 5mC bascule de la double hélice pour être insérée dans la poche catalytique de l'enzyme. Dans cette structure, il apparaît que le groupement méthyle de la cytosine est stabilisé par son insertion entre les groupements hydrophobes d'une alanine et d'une valine. La poche catalytique de l'enzyme comporte suffisamment d'espace pour pouvoir accueillir différents substituants à la position 5 de la cytosine, en accord avec la capacité de l'enzyme de réaliser les oxydations successives

jusqu'à la 5caC. Une information importante concerne la séquence de faible complexité dans le domaine DBSH : celle-ci est orientée vers l'extérieur du site catalytique, ce qui est cohérent avec son rôle potentiel comme domaine d'interaction avec des protéines qui réguleraient l'activité des protéines TET. Cette séquence n'est pas nécessaire à l'activité catalytique de l'enzyme puisque le domaine catalytique de TET1 dans lequel on l'a supprimée conserve toujours son activité d'oxydation des 5mC, de la même manière que l'enzymes TET2 dont on a supprimé une partie de cette insertion^{30,31}.

Tant TET2 que NgTet1 ne comportent pas de domaines CXXC. L'étude des données cristallographiques n'a donc pas pu apporter d'information sur leur rôle et leur interaction éventuelle avec les autres domaines de la protéine. Cependant, d'autres études biochimiques ont pu faire apparaître des rôles apparemment paradoxaux pour ces domaines. Les domaines CXXC sont présents dans de nombreuses protéines interagissant avec l'ADN, mais ils ne possèdent pas tous la capacité à se lier à l'ADN.

Chez la protéine TET1, le domaine CXXC semble moduler l'activité de l'enzyme *In cellulo*³². En effet, dans les cellules HEK293, l'expression de la protéine tronquée TET1 comportant uniquement le domaine catalytique, induit des niveaux de 5hmC plus élevés que l'expression de la séquence complète de la protéine. Cette différence d'activité semble être induite par l'affinité du domaine CXXC pour les CpG non méthylés, puisque TET1 est associée préférentiellement aux îlots CpG peu méthylés et est exclue des îlots CpG hyperméthylés, alors que son domaine catalytique ne présente pas de distribution particulière. La liaison du domaine CXXC de TET1 à des îlots CpG non modifiés a été confirmée par une autre étude, qui montrait aussi que ce domaine était capable de lier *in vitro* et *in vivo* des CpG contenant des 5mC et des 5hmC avec une affinité cependant inférieure à la séquence non modifiée³³. Conflictuellement, une autre étude a rapporté que le domaine CXXC de TET1 se liait très faiblement à l'ADN³⁴.

Chez la protéine TET2, rappelons que CXXC4, ou IDAX, est devenu un gène à part entière lors de l'inversion chromosomique précédemment évoquée au paragraphe 3.1. Cependant, il semblerait qu'une interaction avec TET2 ait subsisté malgré cette séparation³⁵. IDAX préfère les CpG non modifiés aux CpG méthylés dans des expériences *in vitro*. *In cellulo*, IDAX est associée aux régions riches en CpG non méthylées et la co-expression d'IDAX et de TET2 induit une diminution des niveaux de 5hmC ainsi que ceux de TET2 tout en gardant constant les niveaux d'ARNm de TET2. Ces effets sont abolis par la mutation des

résidus d'IDAX impliqués dans la liaison à l'ADN, et par le traitement avec des inhibiteurs des caspases. Les auteurs ont donc proposé un modèle où IDAX va simultanément recruter TET2 sur des séquences spécifiques de l'ADN et activer des caspases qui vont dégrader IDAX et TET2. Ainsi, la localisation et les niveaux de 5hmC pourraient être régulés par ce mécanisme, *via* l'expression d'IDAX.

Le domaine CXXC de TET3, quant à lui, semble avoir uniquement une faible préférence pour les CpG par rapport à d'autres séquences³⁶.

Les domaines CXXC des protéines TET présentent donc un comportement qui pourrait paraître paradoxal : ils n'ont pas d'affinité particulière, voire peu d'affinité, pour les séquences CpG méthylées alors que ces séquences sont la cible des domaines catalytiques de TET. Néanmoins, ils semblent impliqués dans la régulation de l'activité des protéines TET, que ce soit par le recrutement des protéines à des séquences spécifiques de l'ADN ou bien par la médiation de la dégradation de ces protéines, comme on le voit lors de l'interaction entre TET2 et IDAX.

4 Stabilité et mutagénicité de la 5hmC, 5fC et 5caC

Lors de la caractérisation des différents états oxydés de la 5mC, la question de la stabilité et de la mutagénicité des bases 5hmC, 5fC et 5caC, s'est tout naturellement posée et différentes études ont permis d'y répondre. D'une part, il a été montré que la 5hmC, la 5fC et la 5caC sont stables dans des conditions physiologiques³⁷. Par ailleurs, il a été observé par la mesure des températures de fusion que, selon les cas, la présence de 5hmC, soit déstabilise très faiblement, soit ne déstabilise pas la formation de duplexes³⁷⁻³⁹, ce qui diffère donc de l'effet stabilisateur observé pour les duplex comportant la cytosine méthylée⁴⁰. La 5fC et la 5caC n'entraînent pas non plus de déstabilisation de la formation de duplex³⁷. Dans la structure de double hélice de l'ADN, le groupement hydroxyméthyle de la 5hmC est projeté dans le grand sillon de l'ADN, vers l'extrémité 3' du brin d'ADN où elle est présente, et est donc placée à un endroit idéal pour ne pas perturber les interactions stabilisant la double hélice d'ADN³⁸.

L'absence d'effet déstabilisant notoire dans les duplex comportant les dérivés oxydés de la 5mC bat donc en brèche l'hypothèse selon laquelle l'interaction entre les atomes d'oxygène des groupements formyle et carboxyle et l'hydrogène porté par l'azote en position

4 forcerait l'équilibre tautomérique vers la forme imine (Figure 15). Cette observation expérimentale a également été appuyée par les calculs théoriques qui indiquent que la forme amine est la plus stable en solution aqueuse⁴¹.

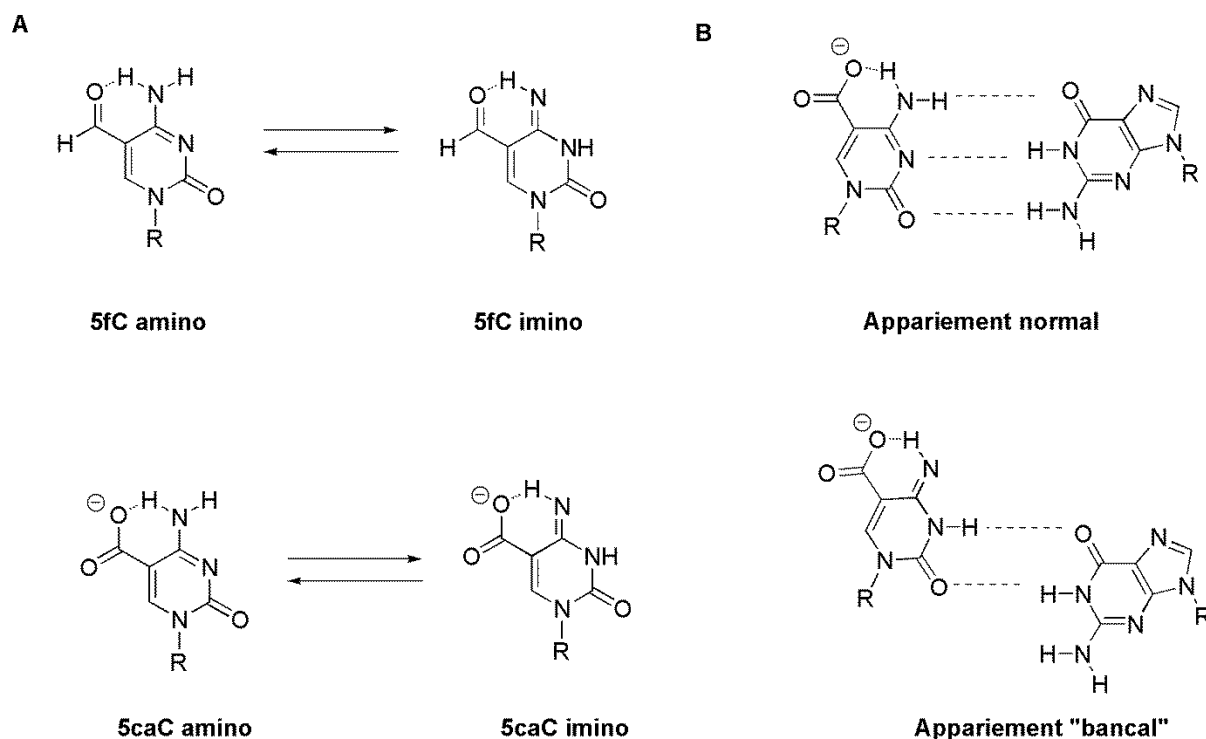


Figure 15. A) Equilibre tautomérique entre les formes amine et imine de la 5fC et de la 5caC. B) Induction d'un appariement « bancal » par le tautomère imine de la 5caC.

De plus, de récents travaux ont montré qu'aucun des dérivés oxydés de la 5mC n'est mutagène^{37,42,43}. Ainsi, la désoxyguanosine est incorporée en face de la 5hmC, de la 5fC et de la 5caC par de nombreuses polymérases de l'ADN y compris bactériennes, qui possèdent pourtant une moins grande fidélité que les polymérases humaines. La vitesse d'incorporation de GTP lors de la lecture de 5fC et de 5caC est réduite, mais ne change pas lors de la lecture de la 5hmC^{43,44}.

L'oxydation de la 5mC par les protéines TET aboutit à la formation de trois nouvelles cytosines modifiées : la 5hmC, la 5fC et la 5caC. Ces trois bases sont stables, ne déstabilisent pas la structure de la double hélice d'ADN et sont correctement reconnues par les polymérases de l'ADN.

5 Distribution des dérivés oxydés de la 5mC chez le mammifère dans l'espace et dans le temps

Les deux études initiales des laboratoires d'Anjana Rao et de Nathaniel Heintz ont employé la chromatographie sur couche mince ainsi que la spectrométrie de masse pour quantifier et identifier, respectivement, la présence de 5hmC. Les estimations de la quantité de 5hmC présentes dans les échantillons allaient de 0.6% des cytosines totales dans les cellules de Purkinje (ou 40% des 5mC) à 0.03% (7% des 5mC) dans les cellules souches embryonnaires de souris. Depuis ces premières publications, de nombreuses techniques ont été développées pour permettre d'une part une quantification précise de la quantité de 5hmC présentes dans des échantillons d'ADN et d'autre part, le séquençage à la base près de la position des cytosines oxydées dans le génome. Ces méthodes seront abordées plus tard, dans l'introduction du chapitre II, mais leurs résultats seront présentés dans les paragraphes suivants.

5.1 Les niveaux de 5hmC varient en fonction de la nature des tissus

Les nombreuses études s'intéressant à la 5hmC ont permis de faire apparaître une différence entre son incidence et celle de la 5mC dans différents types de tissus, qu'ils soient humains ou murins. En effet, alors que les niveaux globaux de 5-méthylcytosine restent constants, les niveaux globaux de 5hmC varient énormément selon la nature du tissu^{45,46}. Comme rapporté dans les premières publications, les niveaux les plus importants de 5hmC sont trouvés dans les différents types de cellules constituant le cerveau mais aussi dans les constituants du système nerveux central, comme la moelle épinière (Figure 16). Des niveaux intermédiaires sont observés dans les poumons, les reins, la vessie et le cœur, tandis que les niveaux les plus faibles se retrouvent dans le foie, la rate, le sang et les testicules.

Figure 16. Quantification par spectrométrie de masse de la 5hmC (A) et de la 5mC (B) dans différents tissus de souris⁴⁶.

Dans le cerveau même, la quantité de 5hmC varie en fonction de la zone concernée^{46,47}. De la même façon que pour les différents types de tissus, les niveaux de 5mC restent constants alors que ceux de 5hmC peuvent ainsi varier du simple au triple, les niveaux les plus élevés étant observés dans l'hypothalamus, le cortex cérébral et l'hippocampe (Figure 17).

Figure 17. Variation des niveaux de 5mC et de 5hmC dans le cerveau de souris. A) Coupe sagittale d'un cerveau de souris B) Niveaux de 5hmC et de 5mC dans les différentes régions du cerveau de souris⁴⁷.

Par ailleurs, à tissu équivalent, les niveaux de 5hmC sont réduits de manière substantielle dans la quasi-totalité des types de cancer solide⁴⁸⁻⁵². Cette diminution des niveaux de 5hmC est accompagnée d'une faible expression des protéines TET et est plus importante que celle des niveaux de 5mC, et ne peut donc pas s'expliquer uniquement par la diminution de cette dernière⁵².

5.2 La 5fC et la 5caC sont présentes en très faible quantité dans l'ADN

Les niveaux de 5fC et de 5caC sont bien plus faibles que ceux de 5hmC dans les cellules de mammifères. Ainsi, dans les cellules souches embryonnaires de souris, la 5fC constitue 0.02% des cytosines totales, et la 5caC 0.003% des cytosines totales^{25,26,53}, ce qui place la 5fC et la 5caC à un et deux ordres de grandeur, respectivement, en dessous de la 5hmC dans les cellules embryonnaires. La purification d'homologues des enzymes TET chez *Coprinosia cinerea* et *Naegleria gruberi* et l'étude *in vitro* de la cinétique des réactions successives d'oxydation a permis récemment d'apporter un élément de réponse sur ces faibles quantités de 5fC et de 5caC. En effet, avec ces deux enzymes, la réaction la plus rapide est celle de l'oxydation de la 5mC en 5hmC, les oxydations suivantes étant plus lentes^{31,54}. Cette différence de vitesse d'oxydation pourrait ainsi expliquer en partie les différentes

significatives dans les quantités de 5hmC, de 5fC et de 5caC présentes dans les cellules souches embryonnaires. Ce faible taux de modification peut aussi s'expliquer par des mécanismes de prise en charge de la 5fC et de la 5caC qui seront présentés dans les paragraphes suivants.

5.3 Cartographie des dérivés oxydés de la 5mC dans le génome

Depuis 2009, de nombreuses méthodes ont été développées pour permettre d'établir une cartographie de plus en plus précise de la présence de la 5hmC, de la 5fC et de la 5caC dans le génome de différents types cellulaires. Les premières études ont employé l'immunoprécipitation, qui permet de déterminer par enrichissement des séquences de 200-1000 bases contenant un nombre indéfini de 5hmC. Des méthodes permettant le séquençage à la résolution de la base sont maintenant disponibles, et ont permis d'obtenir la position exacte des 5hmC dans le génome de différents types cellulaires.

La 5hmC est particulièrement présente dans les séquences intergéniques dans la quasi-totalité des cellules étudiées⁵⁵⁻⁵⁷. La signification de la présence de la 5hmC dans les séquences intergéniques n'est pas encore bien comprise. Certaines études ont montré qu'il y avait une corrélation entre présence de 5hmC dans ces séquences et l'expression des gènes, notamment dans le cerveau, le foie et les cellules embryonnaires de souris^{55,58-60}.

Dans les cellules souches embryonnaires et neuronales de souris, la 5hmC est enrichie dans l'euchromatine, la forme d'ADN la moins compacte, tandis que la 5mC est enrichie dans l'hétérochromatine, la forme la plus compacte^{55,56}. Par ailleurs, il a été montré que les niveaux de 5hmC sur le chromosome X sont faibles⁵⁵. Dans le cerveau, la 5hmC est présente aux promoteurs et dans les régions intragéniques. Il y a peu de recouvrement entre la 5mC et la 5hmC, et on peut observer une corrélation entre la présence de 5hmC dans les séquences intragéniques et l'expression des gènes en question⁵⁹.

Le séquençage de la 5hmC à la résolution de la base a été effectué dans des cellules souches embryonnaires murines et humaines^{61,62}. La 5hmC est présente à proximité des promoteurs et des amplificateurs, mais pas sur les séquences des promoteurs (ou amplificateurs) mêmes. La 5hmC est de plus présente sur des îlots CpG de faible densité. Le

séquençage simultané de 5mC, 5hmC et de 5fC à la résolution de la base dans des cellules souches embryonnaires de souris a permis de nuancer la faible quantité de 5fC présente dans ce type de cellules⁶³. En effet, si le nombre global de sites contenant des 5fC est bien plus faible que ceux contenant des 5mC et des 5hmC, le taux de modification moyen pour un CpG donné est de 22.8%, soit un taux proche de ceux de la méthylation et de l'hydroxyméthylation (25.4% et 17.9%, respectivement). De plus, les 5hmC et les 5fC ont tendance à être présentes dans un contexte d'hémi-modification, au contraire de la 5mC qui était globalement présente symétriquement sur les sites identifiés.

5.4 Variation des niveaux globaux de 5hmC dans le temps

• Variation en fonction de l'âge des tissus

Dans le cerveau, la formation de 5hmC ne se déroule pas en une seule fois mais continue au cours du temps, même après la fin du développement embryonnaire et la naissance. Ainsi, dans le cervelet des souris, la quantité de 5hmC augmente au moins jusqu'à l'âge d'un an⁵⁵. Cette augmentation semble être différente selon la nature des tissus, puisqu'on observe dans l'hippocampe de ces mêmes souris que l'augmentation de 5hmC s'arrête au bout de six semaines (Figure 18).

Figure 18. Evolution des niveaux de 5hmC chez les souris au cours du temps. **A)** Dans le cervelet **B)** Dans l'hippocampe⁵⁵.

Dans ces études menées sur l'évolution des niveaux de 5hmC au cours du temps, il faut souligner que le taux de 8-oxo-guanine, une base issue d'un dommage oxydatif de la guanine, est pris pour contrôle afin de vérifier que les variations des taux de 5hmC observés sont bien liées à l'activité de TET.

• Variation dans les cultures primaires au fil des passages cellulaires

De plus, il a été montré que les niveaux de 5hmC diminuent dans le temps dans les cultures primaires (culture de cellules non immortelles, issues de tissus) au fil des passages cellulaires. On peut observer ce phénomène dans la figure 19, qui présente la détection par

immunofluorescence de 5hmC et de 5mC dans un tissu mammaire sain et dans les cellules mises en culture issues de celui-ci. Les niveaux de 5mC, mesurés par immunofluorescence, ne changent pas après la mise en culture cellulaire alors qu'on observe une diminution du taux de 5hmC au cours du temps, pour arriver à des niveaux négligeables au bout de 28 jours (Figure 19, p4)⁴⁵. Cette diminution est accompagnée d'une réduction de l'expression des enzymes TET. L'adaptation à la culture cellulaire a donc une incidence sur la présence de la 5hmC et l'activité des enzymes TET. Enfin, on observe aussi des niveaux de 5hmC réduits dans certains tissus se divisant très rapidement, comme les cellules souches de l'épithélium intestinal⁵².

Figure 19. Evolution des niveaux de 5hmC et de 5mC lors de la mise en culture de cellules issues de tissus mammaires non tumoraux. p0 correspond au début de la culture, p1 au 7^e jour de culture, p2 au 14^e jour de culture et ainsi de suite. La dernière ligne (α -SS) correspond au contrôle (anticorps contre l'ADN simple brin)⁴⁵.

• Cinétique d'hydroxyméthylation lors du cycle cellulaire

Malgré le grand intérêt pour la 5hmC, très peu d'étude se sont intéressées à l'élément temporel de sa formation lors du cycle cellulaire. La maintenance de la méthylation a lieu pendant la réplication de l'ADN, l'enzyme DNMT1 permettant de conserver les CpG totalement méthylés. Le tempo de l'hydroxylation de l'ADN par rapport à celui de la méthylation est encore peu connu, mais une étude récente du groupe de Shankar Balasubramanian a permis d'apporter des premiers éléments de réponses à cette question fondamentale⁶⁴. L'utilisation de différents marquages isotopiques couplés à une analyse par spectrométrie de masse haute résolution a permis de s'intéresser aux dynamiques d'apparition de la 5mC et de la 5hmC dans des cellules souches embryonnaires de souris (mES) et dans des lignées cellulaires cancéreuses (HCT116 et MCF7). Le principe de ces expériences est le suivant: les cultures sont réalisées en présence du cofacteur donneur de méthyle AdoMet marqué isotopiquement (groupement ¹³C-D₃ à la place de son groupement méthyle). Ceci permet de suivre l'incorporation progressive de ce marquage isotopique dans la population de cellules en division (Figure 20).

Grâce à cette technique, les auteurs ont d'abord simplement mesuré les niveaux de 5mC et de 5hmC à différentes étapes du cycle cellulaire, et ont observé que les quantités des deux modifications de la cytosine restaient constantes tout au long de celui-ci. Cela indique que l'oxydation des 5mC en 5hmC a lieu pendant la réplication de l'ADN, puisqu'on devrait sinon observer une diminution par deux du niveau de celle-ci, la quantité globale d'ADN doublant à cette étape.

Figure 20. Cinétique de la formation de 5hmC (rouge) à partir de 5mC nouvellement formée (bleue) dans des cellules cancéreuses (HCT116 et MCF7) ou des cellules souches embryonnaires de souris (mES). L'ordonnée représente le pourcentage de 5mC ou de 5hmC marquée isotopiquement par rapport à la quantité totale de 5mC ou de 5hmC, respectivement.

Au cours des divisions cellulaires, de plus en plus de marquage isotopique est introduit dans l'ensemble des 5mC. Cependant, ce marquage isotopique n'est pas immédiatement transmis *via* la formation de 5hmC. Ainsi la formation de 5hmC sur le brin synthétisé pendant la réplication de l'ADN s'effectue avec un décalage compris entre 2 et 4 heures par rapport à la formation de 5mC par la maintenance de la méthylation. Une dernière expérience a permis d'évaluer plus précisément cette cinétique de formation de la 5hmC. Cette fois-ci, les cellules ont été placées en présence du milieu marqué isotopiquement pendant deux heures, avant d'être remises dans un milieu standard. Le suivi de l'incorporation par spectrométrie de masse a permis de confirmer la présence du décalage entre la formation de 5mC et de 5hmC, mais aussi d'observer que la majorité de la formation de 5hmC à partir de la 5mC nouvellement synthétisée avait lieu 10 à 25 heures après la méthylation.

Non seulement cette étude permet d'apporter des informations très précieuses sur la cinétique de l'hydroxyméthylation, mais elle a également permis de proposer un modèle pour expliquer la variation des niveaux de 5hmC observée en fonction de la nature des tissus. En effet, si la 5hmC se forme en majorité 10 à 25 heures après la réplication de l'ADN, alors la vitesse de croissance des cellules va influencer directement la quantité de 5hmC formée, au contraire de celle de 5mC qui est formée directement pendant la réplication. Ainsi, les cellules qui vont se diviser les plus rapidement (notamment les cellules cancéreuses, mais aussi les cellules souches de l'épithélium intestinal) vont répliquer leur ADN trop vite pour que la majorité de la 5hmC ait le temps de se former, et donc avoir des taux de 5hmC bas.

6. Les enzymes TET et la 5hmC chez d'autres organismes

La majorité des études se sont concentrées sur l'étude des protéines TET et des produits d'oxydation de la 5mC chez l'homme et les souris, dans des cellules différenciées ou non. Cependant, la présence de 5hmC et sa formation par des homologues des protéines TET a aussi été mise en évidence chez d'autres mammifères et dans d'autres organismes comme les insectes, les poissons, ou encore les amibes. Le tableau 1 récapitule les différents organismes qui ont été étudiés jusqu'à présent.

Organisme	Présence 5hmC	Protéine TET ou homologue caractérisé	Référence
Poulet, poisson zèbre	Chez l'embryon : non	Oui	65,66
	Chez l'adulte : oui		
Amphibiens	Oui	Oui	65
Eucaryotes unicellulaires	Oui	Oui	31,67
Champignon	Oui	Oui	54
Insectes	Oui	Oui	68

Tableau 1. Présence de la 5hmC et des protéines TET dans différents organismes.

En revanche, chez la plante *Arabidopsis thaliana*, la présence de 5hmC a été détectée mais à des niveaux indiquant qu'elle résulte probablement plutôt de dégâts oxydatifs de l'ADN⁶⁹. Cette observation est cohérente avec l'absence d'homologues de TET chez *Arabidopsis* et la majorité des plantes.

La formation de 5hmC par les protéines TET est donc un mécanisme commun à de nombreux organismes très variés, des mammifères aux eucaryotes unicellulaires. Les quantités de 5hmC, et des bases issues des oxydations itératives de la 5mC, la 5fC et la 5caC, varient en fonction des tissus et au cours du développement des organismes. Ces trois nouvelles bases modifiées sont présentes dans des quantités certes faibles, mais non négligeables dans un grand nombre de tissus cellulaires. Leur répartition non uniforme dans les différents tissus ainsi que dans les séquences d'ADN indique que ces oxydations itératives

des 5mC répondent à un ou plusieurs rôles biologiques. Les paragraphes suivants seront consacrés à l'état des connaissances sur les deux rôles biologiques de l'oxydation des 5mC identifiés à l'heure actuelle, ainsi qu'à la description des différents mécanismes où son implication a pu être démontrée.

7. Rôles biologiques de l'oxydation des 5mC par les protéines TET

Concernant le rôle biologique des dérivés oxydés de la 5mC, deux principaux rôles ont émergé dès la caractérisation de ces bases de l'ADN et ont été validés au cours des études ultérieures. Premièrement, l'oxydation des 5mC par les protéines TET peut constituer la première étape dans différents mécanismes de déméthylation active de l'ADN, permettant d'assurer la réversibilité de la 5mC. Deuxièmement, les trois nouvelles bases identifiées peuvent elles aussi avoir un rôle de modification épigénétique, que ce soit en abolissant l'effet répresseur de la méthylation des promoteurs de gènes, ou bien en recrutant des protéines spécifiques. Nous allons maintenant résumer les études qui ont permis de mieux comprendre ces rôles.

7.1 La 5hmC comme intermédiaire dans un processus de déméthylation active de la 5mC

Trois types de mécanismes peuvent être mis en œuvre dans la déméthylation active de l'ADN : la facilitation de la déméthylation passive de l'ADN par les oxydations successives des 5mC, la déméthylation par les mécanismes de réparation de l'ADN, et enfin l'action d'enzymes sur les dérivés oxydés de la cytosine pour catalyser la formation d'une cytosine non modifiée.

7.1.1 Interférence dans la maintenance de la méthylation de l'ADN

Nous avons vu précédemment que DNMT1 était chargée de la maintenance de la méthylation des cytosines dans l'ADN. Cette maintenance se fait *via* l'interaction avec la protéine Ubiquitin-like, containing PHD and RING finger domains, 1, aussi appelée UHRF1, qui va se lier à l'ADN et recruter spécifiquement DNMT1. Il a été montré que l'activité de méthylation de DNMT1 vis-à-vis d'un substrat hémihydroxyméthylé est réduite de plus de 90% par rapport à un substrat hémiméthylé⁷⁰, ce qui situe cette activité à des niveaux similaires à la méthylation d'un motif CpG ne comportant aucune cytosine méthylée^{71,72}. Au contraire, l'activité des DNMT3, enzymes de méthylation *de novo*, ne change pas selon la nature du substrat (non modifié, héli-méthylé ou héli-hydroxyméthylé). L'affinité relative d'UHRF1 pour les CpG héli-hydroxyméthylés ou héli-méthylés qui pourrait expliquer la différence de recrutement de DNMT1 a été étudiée mais a donné lieu à des résultats contradictoires^{73,74}. Enfin, la 5fC et la 5caC entraînent aussi une diminution de l'activité de méthylation DNMT1 non seulement vis-à-vis du CpG comprenant une de ces cytosines modifiées mais aussi vis-à-vis des CpG présents à proximité de celui-ci⁷⁵.

L'oxydation des 5mC par les protéines TET peut donc interférer avec la maintenance de la méthylation par DNMT1, qui ne peut plus reconnaître les CpG oxydés. En l'absence de maintenance de la méthylation sur les duplex comportant des CpG héli-hydroxyméthylés créés lors de la réplication de l'ADN, la méthylation présente à des positions spécifiques va se diluer progressivement, au fur et à mesure des divisions cellulaires et des cycles de réplication de l'ADN (Figure 21).

Figure 21. Facilitation de la déméthylation passive par l'hydroxylation des CpG méthylés. Tiré de⁷⁶.

Dans le cas des DNMT3, l'absence de discrimination entre les CpG méthylés ou hydroxyméthylés par ces enzymes pourrait permettre d'envisager que lors d'événements massifs d'oxydation des 5mC, les DNMT3 puissent prendre le relai de DNMT1 et d'UHRF1 et préserver ainsi une partie de la méthylation de maintenance⁷⁷.

7.1.2 Déméthylation par les voies de réparation de l'ADN

Les dommages oxydatifs, ou bien des défauts dans la réplication de l'ADN, peuvent entraîner l'apparition de mésappariements dans la double hélice d'ADN. Afin de corriger ces défauts qui peuvent entraîner des mutations, il existe des enzymes, les glycosylases, capables de reconnaître ces mésappariements et d'exciser les bases modifiées pour créer des sites abasiques. Ces sites abasiques sont ensuite pris en charge par les systèmes de réparation de l'ADN de type Base excision repair (BER) qui vont ajouter la base complémentaire à celle présente, intacte, sur l'autre brin du duplex d'ADN. Lors de la redécouverte de la 5hmC, une des hypothèses quant à son rôle biologique était alors que les oxydations successives en 5hmC, 5fC et 5caC allaient entraîner la formation de paires 5hmC:G, 5fC:G et 5caC:G qui pourraient être reconnues et prises en charge par ces voies de réparation de l'ADN. Deux mécanismes de ce type, impliquant des glycosylases et le système BER, ont été mis en évidence depuis 2009.

Le premier mécanisme met en jeu la Thymine DNA glycosylase (TDG), une enzyme capable d'exciser les mésappariements T:G. La TDG est capable d'exciser la 5fC et la 5caC dans le contexte des appariements 5fC:G et 5caC:G, et de donner lieu à la formation d'un site abasique, qui va pouvoir être pris en charge par le système BER afin de remplacer ce site abasique par une cytosine (Figure 22)^{26,78}.

Figure 22. Déméthylation de l'ADN par oxydation des 5mC par les protéines TET suivie d'une excision des 5fC ou 5caC par l'enzyme TDG qui va former un site abasique. Celui-ci est ensuite pris en charge par le système BER de réparation de l'ADN. Adapté de ⁷⁷.

Le site catalytique de l'enzyme étant capable d'accepter tous les dérivés oxydés de la cytosine et d'établir des interactions avec eux⁷⁹, la différence de sélectivité pour ces bases peut s'expliquer par une différence d'effets électroniques entre les différents substituants en position 5 qui facilitent plus ou moins la rupture de la liaison *N*-glycosidique par la TDG. L'hypothèse que la 5caC et la 5fC pourraient être sous leur forme tautomère imine (Figure 15), qui déstabiliserait l'appariement avec la guanine et induirait la reconnaissance par la TDG, n'est pas confortée par les calculs théoriques ni par les mesures de stabilité des duplex^{37,41}. D'après ces travaux où des expériences ont été réalisées à différents pH et des

études de mutagenèse dirigée ont été effectuées, l'activité vis-à-vis de la 5caC impliquerait de plus la protonation de l'azote en position 3 de la cytosine.

L'existence de la deuxième voie impliquant des mécanismes de réparation est sujette à débat. Le groupe de Hongjun Song a montré en 2011 que les enzymes de la famille AID/APOBEC (Activation-induced deaminase/Apolipoprotein B mRNA-editing enzyme complex) pouvaient induire une déméthylation de l'ADN⁸⁰. Dans cette étude, la surexpression simultanée de TET1 et d'AID dans des cellules embryonnaires HEK293 donne lieu à la diminution des niveaux de 5hmC par rapport aux cellules où seule TET1 est surexprimée. La surexpression de TET1 dans ces mêmes cellules entraîne l'apparition de 5hmU dont les niveaux diminuent quand TET1 est coexprimée avec l'enzyme SMUG1, une glycosylase spécifique des mésappariements 5hmU:G⁸¹. Cela a conduit les auteurs de cet article à proposer un modèle de mécanisme de déméthylation active des 5mC induite par les enzymes TET où les enzymes de la famille AID/APOBEC vont dans un premier temps désaminer la 5hmC en 5hmU. Dans un deuxième temps les mésappariements 5hmU:G vont être pris en charge par les enzymes de réparation de l'ADN et permettre de revenir à la cytosine non modifiée (Figure 23).

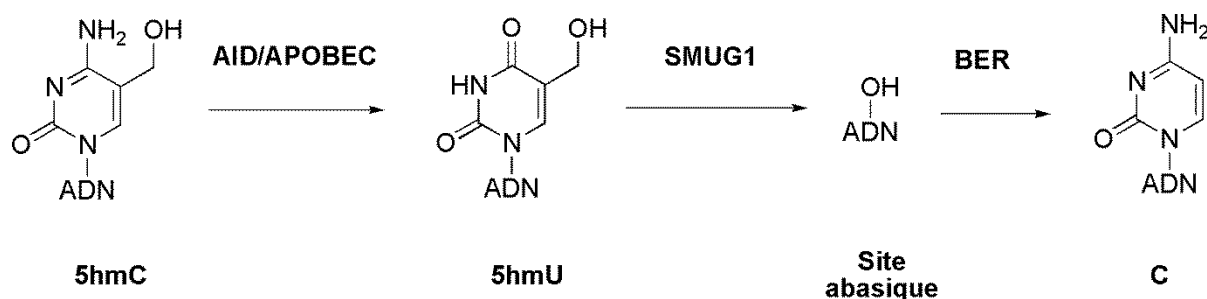


Figure 23. Désamination de la 5hmC en 5hmU et excision par une glycosylase spécifique, SMUG1, suivie de la réparation du site abasique obtenu par le système BER.

Cependant, les résultats de différents groupes de recherches sont venus contredire ce modèle. Ainsi, une étude a montré que l'activité désaminase des enzymes de la famille AID/AIPOBEC dépendait de l'encombrement stérique à la position C5 de la cytosine⁸². En effet, l'activité est réduite d'un facteur 10 entre la C et la 5mC, et n'est pas détectable pour la désamination de la 5hmC en 5hmU. Au contraire, l'expression de la TDG dans des cellules modèles permet de réduire les niveaux de 5fC et de 5caC sans toucher aux niveaux de 5mC et

de 5hmC. Cette observation vient contredire directement la publication précédente. De plus, AID semble agir préférentiellement sur le simple brin, ce qui limiterait son implication biologique. Le groupe de Thomas Carell a aussi apporté des arguments contre cette voie de déméthylation, en utilisant un marquage isotopique des différents nucléosides de l'ADN et une analyse précise par LC-MS/MS²⁹. Dans des modèles de cellules souches embryonnaires, il a montré que la majorité des 5hmU détectées étaient quasi exclusivement formées par l'action des enzymes TET sur la thymidine, capables d'oxyder la base T pour former des mésappariements 5hmU:A. Uniquement 7% des 5hmU présentes dans l'ADN provenaient de la 5hmC, et représentaient 0.06% des niveaux totaux de 5hmC. L'étendue réelle de la désamination des 5hmC par les enzymes AID/APOBEC *in vivo* n'est donc pas encore clairement établie, puisque certaines publications semblent malgré tout établir des liens entre les enzymes AID/APOBEC et des événements de déméthylation^{83,84}.

Chez *Arabidopsis thaliana*, la famille des enzymes DEMETER est capable de reconnaître les 5mC et 5hmC et d'exercer leur activité catalytique sur ces cytosines modifiées⁸⁵. Ainsi, les enzymes ROS1 et DME vont reconnaître la 5hmC et l'exciser d'un double brin d'ADN, puis induire une coupure dans le brin d'ADN *via* leur activité de lyase. Cette coupure va ensuite être réparée par les enzymes du système BER qui vont introduire une cytosine à la place précédente de la 5hmC. Cependant, aucune enzyme possédant une activité similaire n'a été identifiée chez les mammifères.

7.1.3 Déméthylation enzymatique « directe »

La rupture de la liaison C-C reliant le groupement méthyle au cycle pyrimidique de la 5mC n'est pas une réaction triviale. Les études sur des nucléosides modèles montrent que deux phénomènes peuvent faciliter la rupture de cette liaison: l'oxydation de la 5mC en 5hmC, 5fC puis 5caC, ainsi que l'activation du cycle pyrimidique par l'addition de thiol sur la position 6. La saturation de la liaison C5-C6 du cycle de la cytosine est en effet d'une importance cruciale dans la rupture de la liaison C-C. Lorsqu'on réduit la liaison C5-C6 de l'ester méthylique de la 5caC et qu'on effectue ensuite une réaction de saponification, on observe une réaction de décarboxylation rapide, alors que la 5caC reste stable même lors d'un chauffage à reflux pendant plusieurs heures⁸⁶. Une des possibilités pour activer cette liaison C5-C6 est d'ajouter un thiol à la position C6 du cycle pyrimidique.

On constate bien en effet qu'en milieu acide et en présence de thiols, la 5fC et la 5caC peuvent ainsi subir des réactions de déformylation et de décarboxylation, respectivement, pour former la cytosine^{86,87} (Figure 24). L'état d'oxydation du groupement à la position 5 est alors aussi une force motrice de la rupture de la liaison C-C, puisque la réaction de décarboxylation est 11 fois plus efficace que celle de déformylation. Un mécanisme de déméthylation des 5mC par des oxydations successives jusqu'à la formation de la 5caC puis une réaction de décarboxylation activée par l'addition d'un thiol (qui pourrait être une cystéine catalytique comme dans le cas des DNMT à la position 6 de la cytosine) est donc une hypothèse attrayante pour expliquer la déméthylation active de l'ADN.

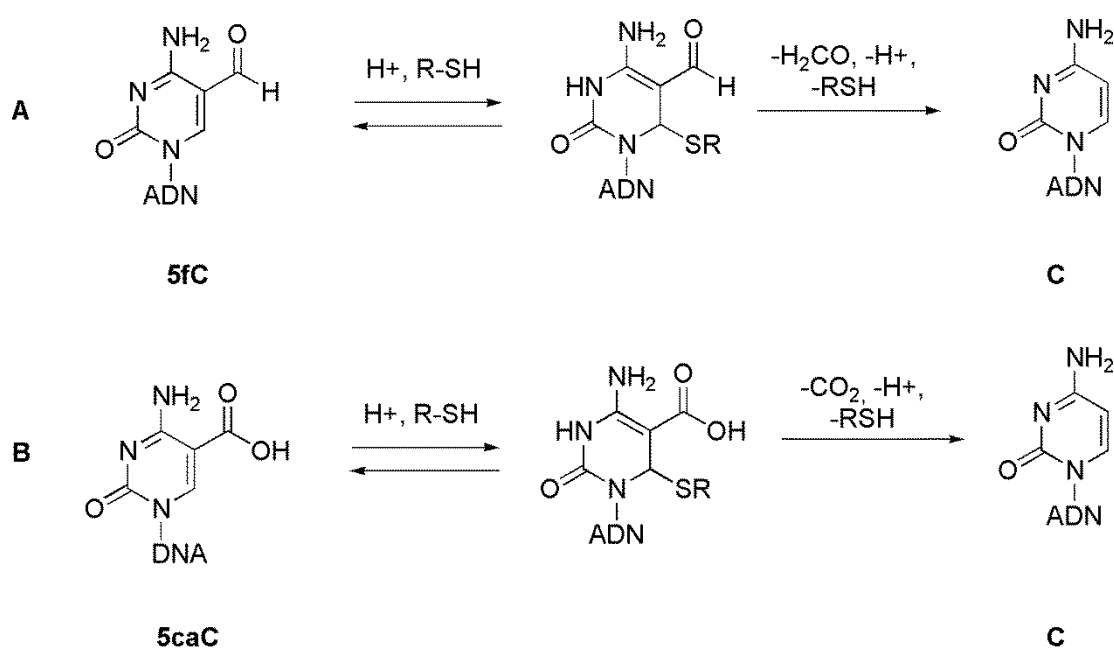


Figure 24. Réactivité en milieu acide et en présence de thiols de **A)** la 5fC **B)** la 5caC.

Cette hypothèse est d'autant plus plausible qu'un mécanisme similaire a déjà été identifié chez les *Fungi*. Chez ces organismes, il existe une voie alternative de récupération de la thymine qui peut se mettre en place lorsque la voie normale de biosynthèse de l'uracile n'est pas disponible⁸⁸. Cette voie de récupération fait intervenir une des premières dioxygénases à Fe (II) et dépendante du 2-OG connues, la Thymine-7-Hydroxylase (T7H). La T7H peut réaliser des oxydations successives à partir de la thymine pour former la 5-hydroxyméthyluracile (5hmU), la 5-formyluracile (5fU) et enfin la 5-carboxyuracile (5caU). La 5caU formée est ensuite décarboxylée par l'action de l'isoorotate décarboxylase pour

aboutir à la formation de l'uracile (U) (Figure 25). De manière remarquable, il a été montré que l'isoorotate décarboxylase peut catalyser la transformation de 5-carboxycytosine en cytosine, bien que moins efficacement que pour son substrat naturel, la 5caU⁸⁹.

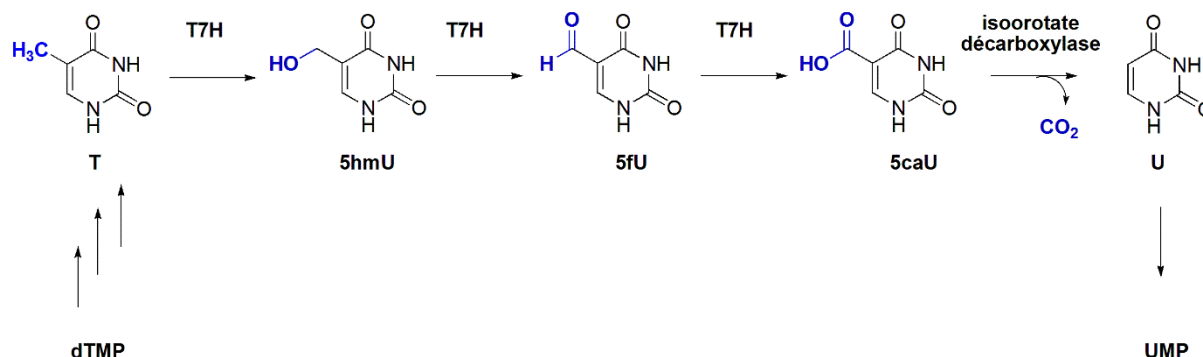


Figure 25. Passage de la thymine à l'uracile, catalysé par la T7H et l'isoorotate décarboxylase.

L'hypothèse d'une voie enzymatique permettant de retourner directement à une cytosine non modifiée, sans passer par un mécanisme d'excision de base, a donc été posée dès la découverte de l'oxydation de la 5mC par les protéines TET. Cependant à ce jour, aucune protéine possédant une activité de 5caC décarboxylase n'a été identifiée. Une étude du groupe de Thomas Carell a cependant permis d'obtenir une preuve indirecte de l'existence d'une activité de décarboxylation dans des extraits cellulaires. Des séquences d'ADN contenant une 5caC marquée isotopiquement sur les azotes N1 et N3 du cycle pyrimidique ont été incubées avec des extraits de cellules souches embryonnaires de souris. L'analyse de ces séquences d'ADN modèle par spectrométrie de masse haute-résolution après l'incubation a permis de déterminer qu'une faible partie des cytosines présentes dans ces séquences contenait le marquage isotopique des 5caC, et d'établir une activité de 5caC décarboxylase dans les extraits de cellules souches embryonnaires utilisés⁸⁶.

Si des études supplémentaires sont nécessaires pour identifier clairement les processus impliqués dans cette décarboxylation, des activités atypiques de déméthylation, déhydroxyméthylation et de décarboxylation réalisées par les ADN méthyltransférases humaines ou bactériennes ont déjà été identifiées (Figure 26). En effet, l'enzyme M.HhaI, une ADN méthyltransférase bactérienne, est capable en l'absence du cofacteur AdoMet de catalyser *in vitro* l'addition de formaldéhyde sur la position 5 d'une cytosine et de former une 5hmC⁹⁰. Cette addition est réversible et la M.HhaI est ensuite capable en l'absence de

formaldéhyde de catalyser le retour à la cytosine non modifiée. Cette étude a d'ailleurs été l'une des premières à suggérer l'idée que la déméthylation des 5mC pourrait se faire *via* une étape d'oxydation du groupement méthyle.

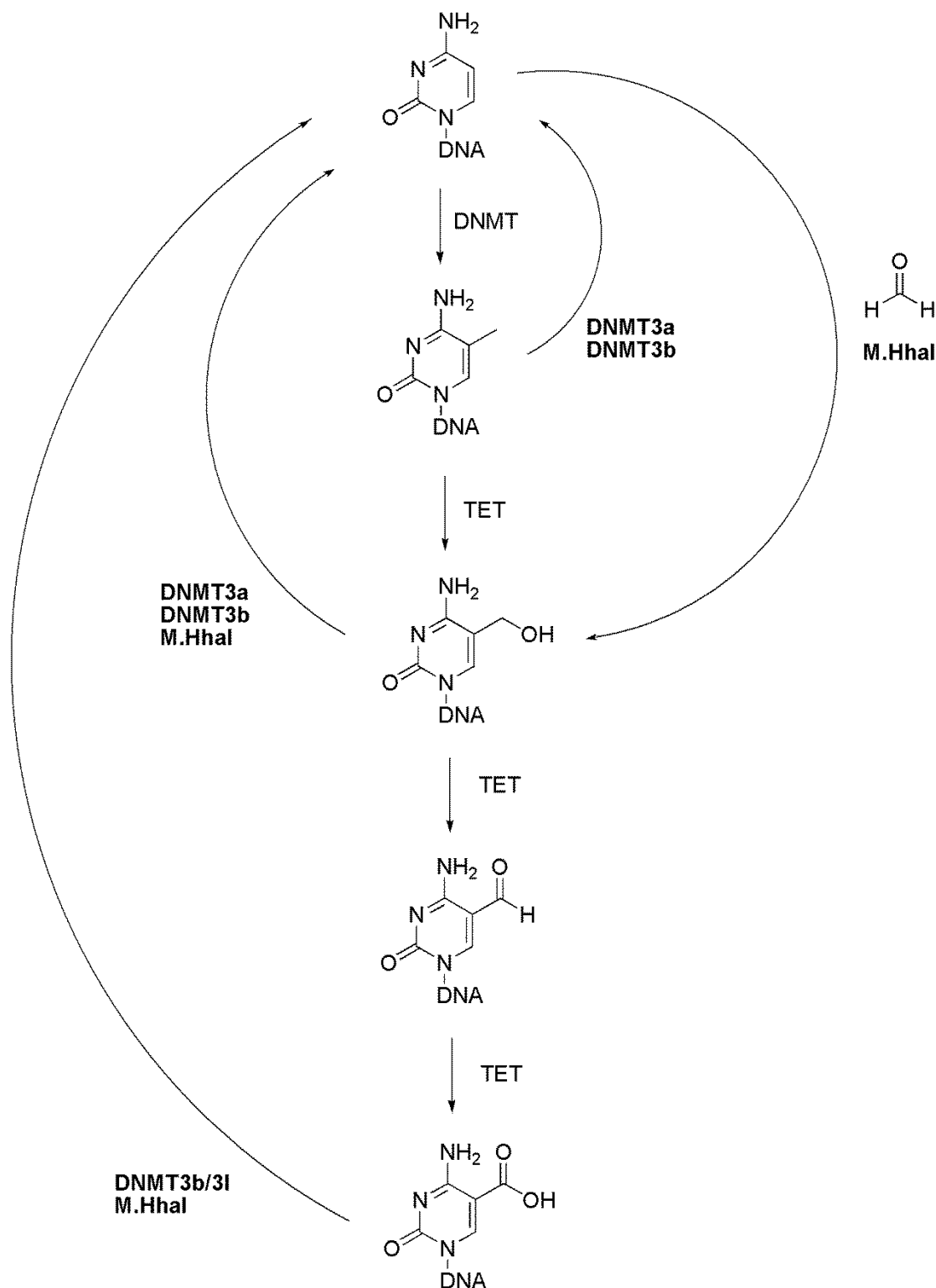


Figure 26. Activités atypiques des DNMT humaines et bactériennes sur la 5mC, la 5hmC et la 5caC observées *in vitro*.

Les DNMT3a et 3b sont aussi capables de catalyser la déshydroxyméthylation de la 5hmC⁹¹. Cette activité est diminuée par la présence de réducteur et augmentée par l'ajout d'oxydant. Le même groupe de recherche a montré que les DNMT1, 3a et 3b possédaient une activité de déméthylase en présence de calcium et que cette activité était aussi sensible à l'addition d'un oxydant ou d'un réducteur⁹². Enfin, l'enzyme M.HhaI peut décarboxyler la 5caC pour revenir à une cytosine non modifiée⁹³. Cette activité est drastiquement augmentée quand on utilise un mutant de l'enzyme possédant un site actif possédant une plus grande cavité. Le complexe DNMT3b/3l possède aussi une activité de ce type, bien que réduite. Contrairement aux publications du groupe de Shen, l'ajout d'un oxydant en solution inhibe totalement l'activité des enzymes. Ces activités de déméthylation, déshydroxyméthylation et décarboxylation, résumées dans la figure 26, sont pour l'instant limitées à des observations *in vitro* dans des conditions éloignées des conditions physiologiques, notamment en absence du cofacteur AdoMet mais suggèrent qu'il y a peut-être une implication d'enzymes ayant un mécanisme proche des DNMT (activation de la position 6 des pyrimidines).

La découverte de l'oxydation des 5mC par les protéines TET a donc permis d'apporter des premiers éléments de réponse à la question fondamentale des mécanismes de déméthylation active de l'ADN. La modification active des 5mC par la famille des enzymes TET entraîne une dilution passive de la méthylation, due au manque de reconnaissance des enzymes chargées de la maintenance de la méthylation de l'ADN. De plus, les oxydations itératives par les enzymes TET de la 5hmC pour donner la 5fC et la 5caC permettent de former des appariements 5fC:G et 5caC:G qui vont être reconnus par une glycosylase, TDG, et entraîner une déméthylation par les mécanismes de réparation de l'ADN. Enfin, l'existence d'un mécanisme de décarboxylation enzymatique menant à la formation d'une cytosine est fortement soupçonnée mais aucune enzyme ne possédant une telle activité n'a été identifiée.

Cependant, de nombreuses observations laissent à penser que la 5hmC, la 5fC et la 5caC ne sont pas uniquement des intermédiaires dans la déméthylation de l'ADN et qu'elles pourraient être des marques épigénétiques à part entière.

7.2 La 5hmC et ses dérivés oxydés : des marqueurs épigénétiques ?

La 5hmC est présente dans des quantités importantes dans le cerveau ainsi que d'autres tissus, jusqu'à 40% des 5mC dans les cellules de Purkinje. De plus, les niveaux de 5hmC augmentent au cours du temps dans différentes régions du cerveau. Ces observations sont en contradiction avec l'hypothèse d'un rôle unique d'intermédiaire dans la déméthylation de l'ADN, qui impliquerait des niveaux faibles de 5hmC dans les différents tissus. De plus, des études récentes ont permis de montrer qu'une quantité importante des 5hmC présente dans l'ADN l'était de manière stable. L'étude de Balasubramanian⁶⁴, présentée précédemment au paragraphe 5.3, a ainsi permis d'en faire une démonstration concluante. Les auteurs ont nourri un couple reproducteur de souris avec un régime spécial dans lequel toute la L-méthionine (le précurseur utilisée dans la biosynthèse de la ADOMET) était remplacée par de la L-méthionine-(¹³C-D₃). Après 4 mois, les niveaux globaux de 5mC et de 5hmC ainsi que l'incorporation du marquage isotopique dans ces deux modifications dans différents tissus du couple reproducteur ont été examinés. Dans tous les tissus du couple reproducteur, on observait que l'incorporation du marquage isotopique dans les 5mC était supérieure à celui dans les 5hmC. De plus, l'incorporation la plus faible du marquage isotopique se retrouvait dans le cerveau, alors que c'est le tissu où la 5hmC était présente dans la quantité la plus importante. Ces expériences suggèrent que la 5hmC présente l'est de manière stable. En effet, si la 5hmC n'était qu'un intermédiaire dans la déméthylation de l'ADN, l'incorporation du marquage isotopique devrait être semblable pour la 5mC et la 5hmC, le système ayant atteint un état stationnaire au bout de 4 mois. La 5hmC est donc présente de manière stable dans différents tissus, et se pose alors la question de savoir si elle participe à la régulation épigénétique, soit en altérant l'effet de la méthylation de l'ADN, soit en recrutant des protéines interagissant spécifiquement avec elle.

7.2.1 Identification de protéines interagissant avec les différentes modifications de la cytosine.

Si l'on veut considérer les effets de l'oxydation des 5mC sur les interactions avec les différentes protéines de liaisons à l'ADN et les facteurs de transcription, trois modèles existent : (i) les bases modifiées 5hmC, 5fC et 5caC peuvent entraîner le recrutement de nouvelles protéines, (ii) elles peuvent perturber l'interaction existant précédemment entre des protéines et les 5mC, (iii) elles peuvent n'avoir aucun impact sur la liaison avec les protéines. Afin d'établir les conséquences de l'hydroxylation des 5mC sur les liaisons avec des protéines, deux voies de recherches ont été explorées : l'étude des interactions de protéines connues pour se lier à l'ADN méthylé avec les autres modifications des cytosines ou la recherche de partenaires protéiques spécifiques par des méthodes de protéomique. Dans le chapitre qui suit, nous allons présenter ce qui a déjà été fait dans ce domaine.

7.2.2. Etudes de quelques protéines de liaison à l'ADN méthylé

• Effets de la 5hmC et de ses dérivés sur la liaison aux facteurs de transcription

De manière générale, les études mesurant les interactions entre les facteurs de transcription et l'ADN comportant les motifs oxydés de la 5mC montrent que les facteurs de transcription possédant une affinité spécifique pour les CpG méthylés ne conservent pas cette affinité pour les 5hmC, 5fC ou 5caC. Ainsi, la liaison des facteurs de transcription HIF1, MAX et USF est diminuée si l'on remplace une 5mC par une 5hmC dans un CpG méthylé, et est complètement abolie si l'on utilise un CpG doublement hydroxyméthylé³⁹. De même, le facteur de transcription Egr1 (Early growth response protein1), impliqué dans la différenciation et la mitogénèse, a une affinité maximum pour sa séquence spécifique lorsque les CpG de celle-ci sont complètement méthylés. Si l'on utilise une séquence comportant des CpG hydroxyméthylés, l'affinité diminue d'un facteur 24 et disparaît totalement si l'on utilise des séquences comportant des 5fC ou des 5caC⁹⁴.

De la même manière, le facteur de transcription Zfp57 (Zinc finger protein 57 homolog), crucial pour un développement normal de l'embryon, a une affinité supérieure pour la 5mC que pour la 5hmC et la 5fC, et ne possède pas d'affinité pour la 5caC⁹⁵. De façon intéressante, des études ont montré que des mutations de ce facteur de transcription peuvent,

dans certains cas, aller jusqu'à inverser la sélectivité pour les cytosines modifiées. Le mutant E182R de Zfp57 possède ainsi une affinité pour la 5caC supérieure à celle pour la 5mC d'un facteur de 10^6 .

Au contraire, le facteur de transcription WT1 (Wilms Tumor protein) reconnaît la 5hmC, la 5fC, la 5mC et la 5caC, respectivement dans un ordre d'affinité croissant, par le biais d'un de ses domaines à doigts de zinc⁹⁴. Les mutations de WT1 interviennent dans certaines formes rares de cancer des reins.

- **Effets de la 5hmC sur la liaison avec les protéines methyl-CpG binding domain (MBD)**

Les protéines MBD sont une famille de protéines contenant un domaine MBD, qui vont se lier préférentiellement aux motifs CpG méthylés et recruter des complexes qui vont remodeler la chromatine, et maintenir une répression transcriptionnelle. Les différentes protéines de cette famille composée des protéines MBD1-4 et MeCP2 sont présentées dans la Figure 27.

Figure 27. Représentation schématique des différentes protéines de la famille MBD et de leurs domaines. D'après ⁹⁷; MBD, methyl-CpG binding domain; TRD transcriptional repression domain; CXXC, domaine CXXC; CC, domaine coiled-coil; G/R, domaine riche en arginines.

MeCP2 a été identifiée comme une protéine reconnaissant tant la 5mC que la 5hmC ⁵⁸. Cette étude montre que MeCP2 est présente en abondance dans le cerveau, et ses niveaux d'expression sont inversement corrélés à la présence de 5hmC. Une des explications de cette corrélation pourrait être que l'interaction de MeCP2 avec les 5mC empêche les protéines TET d'accéder à celles-ci. En effet, il a été observé que la pré-incubation d'un oligonucléotide contenant une 5mC avec un équivalent de MeCP2 permet d'inhiber de 90% la formation de 5hmC par le domaine catalytique de TET1. Cependant la distribution de la 5hmC ne change pas dans des cellules ayant subi un knockout de MeCP2, impliquant que d'autres mécanismes sont à l'œuvre.

Concernant la protéine MBD3, il y a une controverse sur le fait qu'elle puisse reconnaître préférentiellement la 5hmC. Une étude a montré que le complexe MBD3/Nurd

(nucleosome remodeling demethylase, complexe de 7 protéines associées dont MBD3, avec des activités de remodelage de la chromatine et d'histones désacétylases) se liait préférentiellement à la 5hmC plutôt qu'à la 5mC *in vitro* ⁹⁸. Dans cette étude, MBD3 montre une localisation préférentielle dans des régions contenant des 5hmC et co-localise avec TET1 dans des cellules souches embryonnaires. Le knockout de TET1 fait disparaître la localisation de MBD3 sur des séquences riches en 5hmC. Enfin, le knockout de MBD3 diminue l'hydroxyméthylation de manière globale en affectant l'expression de gènes riches en 5hmC, ce qui a conduit les auteurs à attribuer à MBD3 un rôle de médiateur de l'expression. Cependant, ces observations n'ont pas été reproduites par d'autres groupes de recherche^{72,97}. Notamment, l'étude de Baubec et de ses collaborateurs a permis de montrer que si MBD2 et MBD3 se liaient bien à des séquences où étaient présentes les 5hmC ainsi que TET1, cette liaison n'était pas dépendante de la méthylation ou de l'hydroxyméthylation de ces séquences puisqu'on observe toujours la même localisation de ces protéines MBD dans un modèle où la méthylation de l'ADN est empêchée.

Enfin, la protéine MBD4, qui possède un domaine exerçant une activité de glycosylase, peut reconnaître des séquences comportant des CpG hémihydroxyméthylés ou hémiformylés⁷⁴. Cependant, les CpG modifiés sur les deux brins ne sont pas reconnus par cette protéine, et elle ne peut exciser que les mésappariements T:G ou 5hmU:G et pas les 5hmC ni les 5fC ⁷².

La plupart des études semblent donc montrer que l'hydroxyméthylation des CpG méthylés a pour effet d'abolir l'interaction avec les protéines de la famille MBD, ce qui est un argument en faveur d'un rôle d'inversion de la répression transcriptionnelle exercée par les protéines s'associant aux séquences méthylées de l'ADN. MeCP2 semble être une exception à cette règle et semble jouer un rôle important, notamment dans le cerveau.

7.2.4 Recherche de protéines interagissant spécifiquement avec 5hmC et ses dérivés.

Pour tenter d'identifier de nouvelles protéines interagissant avec les cytosines modifiées présentes dans l'ADN, l'une des méthodes employées a été d'effectuer des expériences dites de « pull-down ». Elles consistent à incuber des séquences d'ADN comportant les différentes cytosines dans des extraits cellulaires de différentes natures. Après

l'incubation, des étapes de purification et de concentration permettent de récupérer uniquement les protéines interagissant de manière spécifique avec les séquences d'ADN utilisées. Les protéines récupérées par cette méthode sont ensuite analysées et identifiées par gel d'électrophorèse, par immunofluorescence ou par spectrométrie de masse. La comparaison entre les différentes protéines récupérées selon la nature de la séquence d'ADN utilisée permet d'identifier les différents profils d'interaction des protéines. Souvent, ces résultats sur des extraits cellulaires sont confirmés par l'expression et la purification des protéines recombinantes d'intérêt suivies d'études *in vitro*.

Une expérience de pull-down avec des séquences comportant des 5mC, des 5hmC ou des cytosines non modifiées a ainsi permis d'identifier par électrophorèse ZBTB2, un facteur de transcription impliqué dans la régulation du cycle cellulaire, comme protéine reconnaissant des oligonucléotides contenant la base C ou la 5mC mais pas la 5hmC, dans des extraits de trois lignées cancéreuses⁹⁹. Ces résultats ont été confirmés par la purification de la protéine recombinante et l'étude de son interaction *in vitro*.

Deux études se sont consacrées à la tâche d'identifier, par spectrométrie de masse, les différents profils de protéines interagissant avec les dérivés oxydés de la cytosine. Les groupes de Balasubramanian et de Reik ont incubé des séquences de promoteurs comprenant des cytosines, des 5mC, des 5hmC et des 5fC en présence d'extraits de cellules souches embryonnaires de souris. L'analyse par spectrométrie de masse des différents échantillons de protéines enrichies a d'abord permis de confirmer l'affinité d'Uhrf1 pour les séquences contenant des 5mC ou des 5hmC, ainsi que celle de Td pour les promoteurs contenant des 5fC. Au contraire de l'étude de Yildirim et al,⁹⁸ présentée précédemment, la protéine MBD3 ne présentait pas de liaison spécifique à 5hmC comparée à 5mC, mais plutôt une affinité pour 5mC ou pour 5mC/5fC selon la séquence de promoteur utilisée. La majorité des protéines identifiées se liant préférentiellement à 5fC avaient une fonction dans la régulation de la transcription et de la chromatine, comme par exemple le complexe Nurd. Le knockdown des protéines identifiées comme interagissant préférentiellement avec la 5fC n'a abouti à l'augmentation des niveaux de 5fC et de 5caC que dans l'unique cas de la TDG. Cela permet d'écarter le rôle de ces protéines dans la réparation de l'ADN et l'excision de 5fC/5caC et plutôt de leur attribuer un rôle de reconnaissance de 5fC comme modification épigénétique.

La deuxième étude a été réalisée par les groupes de Thomas Carell et de Michiel Vermeulen, en utilisant des techniques de protéomique quantitative qui permettent de

comparer directement par spectrométrie de masse les quantités de protéines présentes dans les échantillons¹⁰⁰. Ils ont utilisé des séquences comportant des cytosines, 5mC, 5hmC, 5fC et 5caC et les ont incubés dans trois types différents d'extraits cellulaires : les cellules souches embryonnaires de souris, les cellules souches neurales de souris et les cerveaux adultes de souris. Dans des extraits de cellules souches embryonnaires de souris, MeCP2 et Uhrf1 ont été confirmés comme lecteurs des 5mC et des 5hmC, ainsi que Thy28, une protéine non caractérisée liée à l'apoptose. La protéine MBD3 n'est pas un lecteur des 5hmC mais au contraire des cytosines non modifiées. Cette interaction préférentielle pour la cytosine par rapport à la 5hmC a été confirmée avec la protéine recombinante purifiée. Toujours dans les cellules souches embryonnaires de souris, les séquences contenant des 5fC et des 5caC recrutent principalement des protéines impliquées dans la réparation des dégâts oxydatifs, avec peu de recouvrement entre 5fC et 5caC, ce qui suggère des rôles différents pour ces deux modifications de la cytosine. La glycosylase TDG se lie préférentiellement avec la 5fC et la 5caC et pas avec la 5hmC, confirmant une nouvelle fois son rôle dans la déméthylation de l'ADN. La 5fC, au contraire de la 5caC, enrichit spécialement des protéines impliquées dans la réparation de l'ADN, comme p53, un facteur de transcription régulant la mitose et l'apoptose des cellules et dont les mutations sont impliquées dans un grand nombre de cancers. Les lecteurs de la 5caC comprennent des protéines appartenant à des complexes remodeleurs de la chromatine mais aussi DNMT1. Globalement, cette étude montre que la cytosine et la 5mC recrutent beaucoup plus de protéines que la 5hmC, et beaucoup moins que la 5fC et la 5caC. Cette observation, confirmée dans l'étude présentée précédemment, semble orienter la 5fC et la 5caC vers un rôle de modification épigénétique, au contraire de la 5hmC qui serait un intermédiaire dans la déméthylation des cytosines.

Dans les cellules souches neurales de souris, un effet délétère de l'oxydation des 5mC sur la liaison avec les facteurs de transcription a été observé. La majorité des protéines interagissant avec la 5hmC est impliquée dans des processus de réparation de l'ADN. La protéine Uhrf2 est un lecteur spécifique des 5hmC, et son interaction avec la 5hmC et dans une moindre mesure avec la 5fC et 5caC a été vérifiée avec la protéine recombinante. De plus, la co-expression de la protéine avec le domaine catalytique de TET1 dans des cellules HEK293T, des cellules rénales embryonnaires humaines, aboutit à l'augmentation des niveaux de 5hmC, 5fC et 5caC. Cette observation permet de proposer pour Uhrf2 un rôle de facilitateur de l'oxydation successive des 5mC par les protéines TET, peut-être en faisant

basculer la 5hmC en dehors de la double hélice d'ADN et en la rendant plus accessible aux protéines TET.

Enfin, dans les extraits protéiques de cerveaux adultes de souris, et au contraire des cellules souches neurales ou embryonnaires, une plus grande quantité de protéines interagissant avec la 5hmC qu'avec la 5mC a été observée. Un des lecteurs spécifiques de la 5hmC caractérisé est Thap11 (aussi appelée Ronin), une protéine exprimée en grande quantité dans certaines parties du cerveau et notamment les cellules de Purkinje, impliquée dans l'embryogénèse et la pluripotence des cellules souches embryonnaires¹⁰¹.

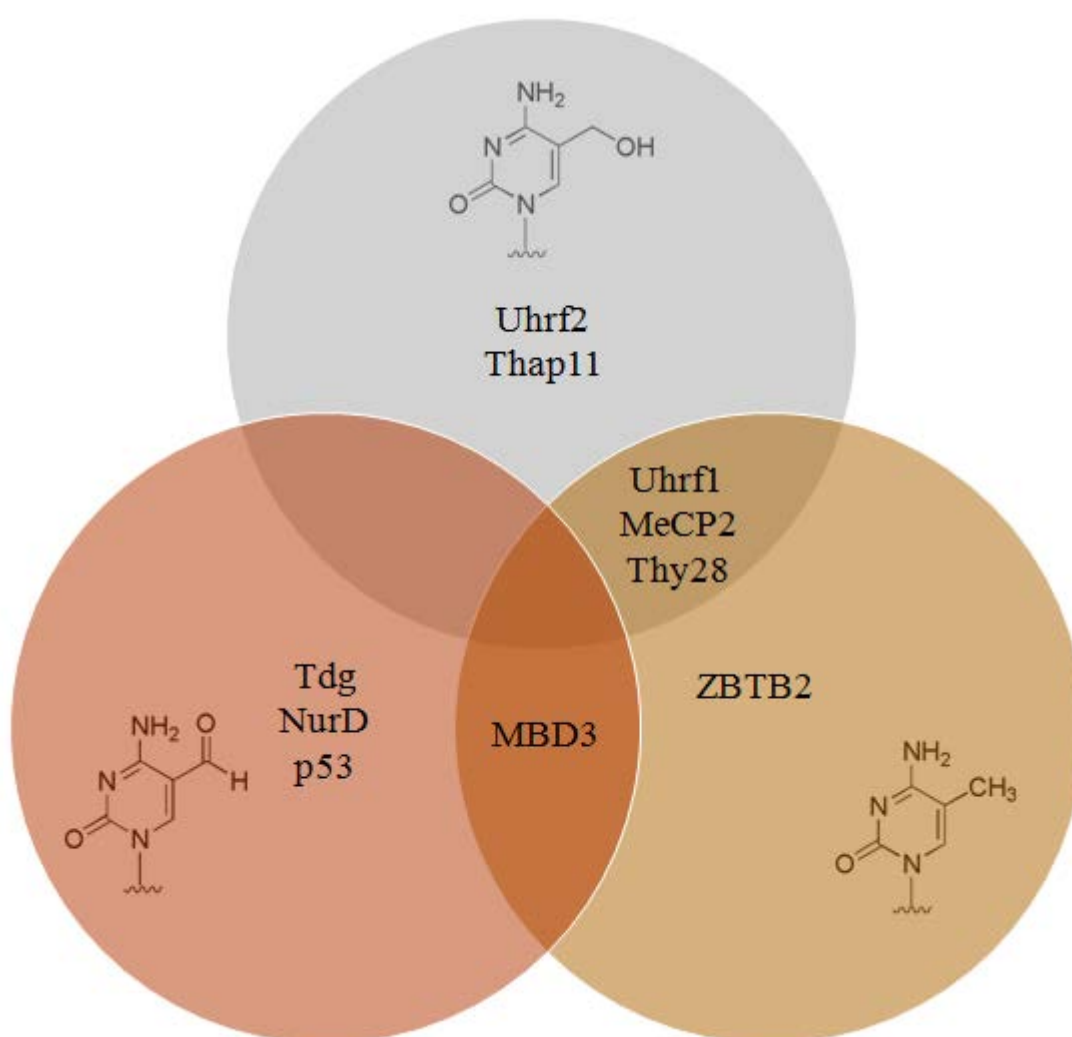


Figure 28. Aperçu des différentes protéines identifiées lors des expériences de pull-down selon leur affinité pour la 5mC, la 5hmC ou la 5fC.

En conclusion, on s'aperçoit que selon la nature du tissu étudié, on observe différents profils de protéines se liant aux bases modifiées de l'ADN. Les interactions entre 5hmC et les protéines sont donc régulées dynamiquement au cours du développement et de la différenciation. Les méthodes employées dans les expériences de protéomique ne permettent cependant pas de distinguer entre les protéines se liant directement aux bases modifiées et celles présentes dans des complexes interagissant avec ces bases. De plus, la nature des liaisons observées peut être dépendante de la séquence d'ADN dans laquelle on introduit les bases modifiées. De nombreuses études sont donc encore nécessaires pour identifier l'ensemble des protéines interagissant avec la 5hmC, la 5fC et la 5caC et les mécanismes qui régissent ces interactions.

8. L'implication de la 5hmC dans des processus biologiques et des pathologies : les dernières avancées

Depuis 2009, une attention très grande a été portée à l'identification de l'implication dans différents phénomènes biologiques des protéines TET et de leur régulation, ainsi que celle des trois nouvelles bases oxydées obtenues à partir de la cytosine. Le paragraphe suivant présentera leur implication dans le développement et la différenciation, ainsi que dans différentes pathologies notamment cancéreuses.

8.1 La 5hmC dans le développement

8.1.1 Variation des taux de 5hmC dans le développement embryonnaire

Pendant le développement embryonnaire des mammifères, des changements drastiques dans le profil de méthylation de l'ADN se déroulent avec des événements de déméthylation et de méthylation *de novo* globaux. Des études récentes se sont intéressées à l'implication des protéines TET et des dérivés oxydés de la cytosine au cours du développement. Ces études ont permis d'éclairer d'autres études plus anciennes portant sur la méthylation pendant

l'embryogénèse et on peut distinguer plusieurs phases selon le stade du développement où les taux de protéines TET et de dérivés de la cytosine sont très différents (Figure 29).

Figure 29. Evolution des niveaux de 5hmC et de 5mC à différentes étapes du développement embryonnaire. **A)** Dans tout l'embryon ; **B)** dans les cellules germinales primordiales²².

- **Déméthylation de l'ADN paternel au stade du zygote.**

Dans le zygote, la cellule-œuf formée par la fertilisation d'un ovocyte par un spermatozoïde, on observe une perte quasi-totale de la méthylation dans l'information génétique paternelle¹⁰² qui ne peut être liée à un phénomène de déméthylation passive puisqu'elle est, à ce stade, indépendante de la réplication. Cette déméthylation globale s'accompagne d'une augmentation des niveaux de 5hmC, 5fC et 5caC^{103,104}. L'utilisation de siRNA dirigée contre TET3 semble indiquer que TET3 est responsable de cet épisode d'oxydation globale des 5mC présentes dans l'information génétique d'origine paternelle. Ensuite, lors du tout début de la division cellulaire, on observe une diminution de 5hmC, 5fC et 5caC qui semble due à une dilution au cours de la réplication et non à un mécanisme actif, par exemple par l'enzyme TDG et les voies de réparation de l'ADN.

Bien que le mécanisme exact n'ait pas encore été élucidé, il est probable que l'oxydation rapide des 5mC dans le zygote permette de faciliter la déméthylation observée, en diminuant l'efficacité de la méthylation de maintenance par DNMT1. Un mécanisme de déméthylation active des 5fC et 5caC par l'enzyme TDG et le système BER, s'il est à l'œuvre, est probablement limité. En effet, la génération de sites abasiques à une très grande ampleur pendant une étape cruciale du développement embryonnaire précoce semble trop dangereuse pour l'intégrité génomique pour constituer l'unique mécanisme à l'origine de la déméthylation dans le zygote.

- **Stade blastula (cellules souches pluripotentes)**

Plus tard dans le développement embryonnaire, lors la transition du stade morula (8-16 cellules quasi identiques) au blastocyste, on observe une forte augmentation de la méthylation

dans la masse cellulaire interne, un groupe de cellules pluripotentes qui vont pouvoir ensuite se différencier pour donner les différents organes de l'embryon. On observe que TET1 et TET2 sont exprimées en grande quantité et semblent avoir un rôle dans la différenciation des cellules souches embryonnaires, similaires aux cellules pluripotentes de la masse interne^{24,56,105}.

- **Déméthylation dans les cellules germinales primordiales**

Entre le 9^e et le 10^e jour du développement embryonnaire, les cellules germinales primordiales subissent un nouvel événement de déméthylation globale (Figure 29b). Cette déméthylation, de manière similaire à celle observée dans le zygote, est accompagnée de la formation de 5hmC par TET1 et TET2¹⁰⁶. La 5mC et la 5hmC disparaissent progressivement, de manière cohérente avec une dilution au cours de la réplication. L'inactivation de la méthylation de maintenance par la conversion des 5mC en 5hmC est accompagnée par la régulation négative d'Uhrf1, partenaire de DNMT1, ainsi que des DNMT3A et B, responsables de la méthylation de novo.

8.1.3 La 5hmC et les protéines TET dans la reprogrammation de cellules souches pluripotentes

Une des grandes avancées dans le domaine des cellules souches embryonnaires est arrivée avec la découverte que l'on pouvait induire la formation de cellules dites cellules souches pluripotentes induites (iPSC), similaires à des cellules souches embryonnaires, à partir de cellules fibroblastiques adultes¹⁰⁷. Cette induction se fait à l'aide de la surexpression d'une combinaison de facteur de transcription abrégée en OSKM (car elle comprend les protéines OCT4, SOX2, KLF4 et MYC). Bien que peu efficace, la possibilité de générer des cellules semblables aux cellules souches embryonnaires offre une énorme opportunité pour de nombreuses applications médicales et permet aussi de contourner les problèmes éthiques liés à l'utilisation de cellules souches embryonnaires humaines. Comme la reprogrammation implique l'activation des facteurs de la pluripotence par une déméthylation de leur promoteurs et amplificateurs¹⁰⁸, la question de l'implication des protéines TET dans ce processus s'est posée. Une étude a montré que l'hydroxylation des 5mC par les protéines TET et l'association de ces dernières avec différentes protéines semblent faciliter la reprogrammation de cellules adultes différenciées en cellules souches pluripotentes induites¹⁰⁹.

8.2. Mécanismes de régulation des protéines TET

Plusieurs études se sont également penchées sur les phénomènes de régulation des protéines TET faisant apparaître une autre couche de régulation de l'hydroxyméthylation *via* (i) des microARN qui perturbent la traduction de l'ARNm de TET en protéine¹¹⁰⁻¹¹³ (ii) une modification post-traductionnelle de TET récemment caractérisée ; ainsi une O-linked β -N-acetylglucosamine transférase (OGT) est capable de modifier les trois enzymes TET^{114,115} sans qu'une influence sur leur activité catalytique n'ait été détectée pour le moment (iii) la dégradation par les calpaïnes, des protéases à cystéine, dont les TET sont des substrats et qui semble aider à réguler finement les niveaux de TET au cours du développement et de la différenciation¹¹⁶.

8.3 La 5hmC et les protéines TET dans le cancer

La première identification de TET1, bien avant que sa fonction de dioxygénase ne soit mise à jour, était celle d'un partenaire de fusion de la protéine MLL (myeloid/lymphoid, or mixed lineage, leukemia) dans certaines leucémies^{117,118}. MLL est une protéine qui méthyle les histones et est un régulateur positif de l'expression des gènes. La translocation du gène de TET1 du chromosome 10 au chromosome 11 crée une protéine de fusion avec MLL, observée dans certaines leucémies myéloïdes aiguës, ce qui a donné son nom aux protéines de la famille TET (ten to eleven translocation protein). Dans ce type de cancer, la protéine TET1 est régulée positivement alors que les protéines TET2 et TET3 ne sont pas affectées, et au contraire de nombreux autres cancers, les niveaux globaux de 5hmC sont augmentés¹¹⁹. Cette surexpression de TET1 est induite par la liaison de protéines de fusions avec MLL avec les promoteurs de TET1. TET1 co-localise avec des gènes cibles des protéines de fusion de MLL, et l'expression de ces gènes est diminuée quand l'expression de TET1 est bloquée. Les auteurs de cette dernière étude proposent donc un modèle où la surexpression de TET1 est induite par les protéines de fusion avec MLL, qui vont ensuite agir de concert avec TET1 pour réguler positivement l'expression de gènes impliqué notamment dans la prolifération cellulaire et l'inhibition de l'apoptose et de la différenciation cellulaire, entraînant l'apparition de la leucémie. Dans les cancers de la prostate et du sein, TET1 est régulée négativement et

cette régulation négative facilite la croissance tumorale et la métastase, et est corrélée avec un pronostic de survie défavorable¹²⁰.

Dans les hémopathies malignes (ou cancers du sang), on observe des mutations fréquentes de la protéine TET2¹²¹⁻¹²³. Les mutations dans TET2 sont présentes ainsi dans plus de 20% des syndromes myélodysplasiques et des leucémies myéloïdes aiguës, ainsi que dans 50% des leucémies myélomonocytaires chroniques. Ces mutations sont accompagnées d'hypohydroxyméthylation^{124,125}, suggérant qu'elles résultent en la perte de la fonction d'hydroxyméthylation de la protéine TET2. Dans des souris où TET2 est inactivée, on observe des niveaux de 5hmC réduits dans la moelle osseuse et la rate¹²⁶ qui s'accompagnent d'une perturbation de la différenciation des cellules sanguines, similaire à celle observée lors des mutations de TET2 sans pourtant engendrer l'apparition d'hémopathies malignes. Cela est cohérent avec les observations sur les patients humains, où les mutations dans TET2 sont souvent accompagnées de mutations dans d'autres gènes¹²⁷. Ainsi, la perte de fonction de TET2 n'est pas suffisante pour initier l'apparition du cancer ce qui peut s'expliquer par le fait que TET1 et TET3 vont pouvoir prendre le relai de TET2 lorsque celle-ci est mutée.

Enfin, les protéines TET peuvent voir également leur activité perturbée indirectement par des pertes de fonction d'autres voies métaboliques et le cas des mutations des isocitrates deshydrogénases (IDH) en est un exemple frappant. Lors d'une des réactions du cycle de Krebs, les isocitrates dehydrogénases (IDH), catalysent l'oxydation et la décarboxylation de l'isocitrate en 2-oxoglutarate, qui rappelons-le est indispensable à l'activité des TET. Des mutations dans les enzymes IDH1 et IDH2 sont trouvées dans plus des trois-quarts des gliomes et glioblastomes ainsi que dans 20% des leucémies myéloïdes aiguës. Certaines mutations ont pour conséquence la perte de fonction normale de l'enzyme qui produit non pas du 2-OG mais un analogue réduit, le (R)-2-hydroxyglutarate (2-HG)¹²⁸, qui est un inhibiteur général des dioxygénases à Fe (II) et 2-OG, et plus spécifiquement des protéines TET¹²⁹. Ceci a pour effet d'entraîner la réduction des niveaux de 5hmC de façon drastique. De manière similaire aux mutations dans TET2, les mutations dans IDH1 et IDH2 ne suffisent pas à entraîner l'apparition des hémopathies malignes^{130,131}. De la même façon, la perte de fonction de deux autres enzymes du cycle de Krebs, la fumarate hydratase (FH) et la succinate déshydrogénase (SDH), a été identifiée dans certains cancers¹³². Cette perte de fonction entraîne l'accumulation de fumarate et de succinate, respectivement, qui peuvent de façon

similaire à 2-HG être des inhibiteurs des dioxygénases à Fe (II) et notamment des enzymes TET.

8.4 La 5hmC dans les maladies neurodéveloppementales et neurodégénératives

Depuis quelques années, on sait que des facteurs épigénétiques sont associés à certaines maladies neurodéveloppementales et neurodégénératives et des études ont donc été réalisées en vue de mieux comprendre le rôle et l'impact de ces modifications de la cytosine sur des pathologies comme la schizophrénie, les troubles du spectre autistique, le syndrome de Rett ou encore la maladie d'Alzheimer.

• Schizophrénie

La schizophrénie touche environ 1% de la population générale et de nombreux facteurs impliquant son apparition ont été mis à jour, tant génétiques que neurobiologiques, neurodéveloppementaux et socio-environnementaux. Cependant, les mécanismes exacts à l'origine de cette maladie psychiatrique ne sont pas encore élucidés. Une étude a récemment montrée que l'expression de TET1 ainsi que la quantité de 5hmC sont plus importantes dans le lobe pariétal inférieur de patients psychotiques (diagnostiqués comme schizophrènes ou atteints de troubles bipolaires)¹³³ et pourrait donc avoir un rôle dans l'apparition de la schizophrénie et des troubles bipolaires.

• Troubles du spectre autistique

Les troubles du spectre autistique sont caractérisés par une perturbation des interactions sociales et de la communication, ainsi que par des comportements stéréotypés et répétitifs¹³⁴. Dans le cervelet de patients autistiques, on détecte une augmentation de l'hydroxyméthylation dans des promoteurs des gènes impliqués dans le développement du cerveau puis ensuite dans le maintien de la plasticité neuronale¹³⁵. Cette augmentation de l'hydroxyméthylation, par rapport à des patients sains a lieu sans changement dans la méthylation et la présence de 5hmC aux promoteurs des gènes corrèle avec la liaison de la protéine MeCP2.

• Le syndrome de Rett

Le syndrome de Rett est une maladie génétique neurologique touchant quasi-exclusivement les femmes, avec une incidence de 1/10 000, et dont les symptômes, compris dans les troubles du spectre autistique, apparaissent 6 à 18 mois après la naissance. Dans 90% des cas, le syndrome de Rett est accompagné d'une mutation dans le gène codant la protéine MeCP2. Une de ces mutations présente la particularité de diminuer l'affinité de la protéine pour la 5hmC sans modifier son affinité pour la 5mC⁵⁸. Le rôle de l'interaction entre MeCP2 et la 5hmC, toutes deux présentes en grande quantité dans le cerveau, dans les mécanismes qui donnent lieu au syndrome de Rett doivent cependant être précisés.

• La maladie d'Alzheimer

La maladie d'Alzheimer est une maladie neurodégénérative caractérisée par un déclin progressif et irréversible des capacités cognitives. C'est la forme la plus fréquente de démence en France et dans le reste du monde. La perte de neurones et de synapses induisant une atrophie des régions affectées est un des symptômes observé dans la maladie d'Alzheimer. Récemment, une étude a rapporté des niveaux accrus de 5mC et de 5hmC dans des lobes du cortex cérébral de patients atteints de la maladie d'Alzheimer¹³⁶. Ces augmentations étaient corrélées avec des marqueurs classiques de la maladie d'Alzheimer, comme les feuillets amyloïdes β et τ , ainsi que les taux d'ubiquitine. Dans un autre lobe du cortex cérébral de patients à un stade précoce ou avancé d'Alzheimer, une augmentation des niveaux de 5hmC, de 5mC et d'expression de la protéine TET1 est accompagnée d'une diminution des niveaux de 5fC et de 5caC, en comparaison avec des patients non atteints de la maladie d'Alzheimer¹³⁷. Aucune étude ne s'est pour l'instant intéressée à la caractérisation de gènes spécifiquement touchée par l'augmentation des niveaux de 5hmC.

Toutes ces observations permettent donc d'esquisser un rôle des perturbations de l'hydroxyméthylation et de la protéine MeCP2 dans les troubles neurodéveloppementaux.

9. Conclusion

La redécouverte de la 5hmC en 2009 a engendré un domaine de recherche extrêmement actif et compétitif, où de nombreuses publications permettent la progression de nos connaissances sur les mécanismes d'hydroxylation de la 5mC par les protéines TET. Les publications originelles des groupes d'Anjana Rao et de Nathaniel Heintz ont été, au jour du 18 décembre 2014, citées respectivement 1251 et 736 fois, et une recherche dans la base de données Web of Science trouve plus de 700 publications sur le sujet, dont la quasi-totalité a été publiée entre 2011 et aujourd'hui (Figure 30). Dans ce domaine en constante évolution, certaines publications peuvent parfois présenter des résultats contradictoires, ou bien même réfuter des hypothèses présentées récemment, comme le montre l'exemple de la déamination des 5hmC par l'enzyme AID. Cependant, un certain consensus commence à émerger sur les fonctions des enzymes TET et les rôles biologiques de l'oxydation enzymatique des 5mC.

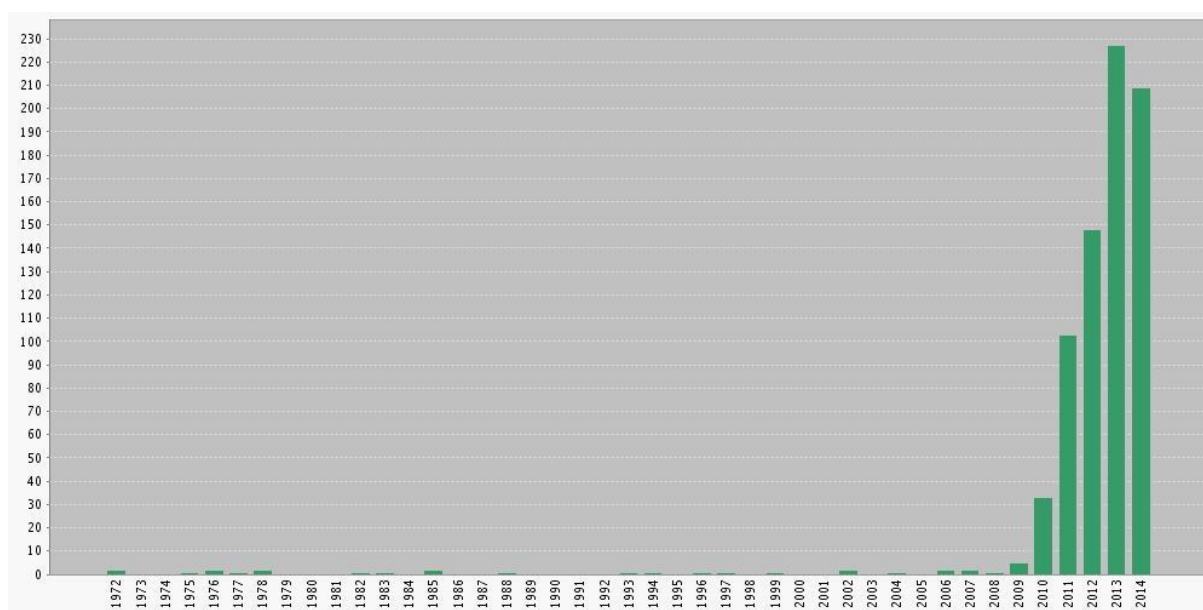


Figure 30. Evolution du nombre de publications dans le domaine de la 5-hydroxyméthylcytosine depuis les années 1970. (Recherche du mot « 5-hydroxyméthylcytosine » dans la catégorie « Topic », base de donnée « Web of Science Core collections »).

Les protéines TET existent dans de nombreux organismes, et peuvent catalyser les oxydations successives de la 5-méthylcytosine en 5-hydroxyméthylcytosine, 5-formylcytosine et 5-carboxycytosine. Ces modifications sont stables et non-mutagènes, et leur rôle biologique commence à être bien défini. L'oxydation de la 5mC permet ainsi d'initier différents

mécanismes de déméthylation active de celle-ci qui permettent un retour à l'état non modifié. De plus, chacune de ces modifications interagit avec des groupes de protéines spécifiques, dont certaines sont impliquées dans la régulation de la chromatine, permettant de commencer à entrevoir un statut de modification épigénétique pour la 5hmC, renforcé par les découvertes qu'une grande partie des 5hmC présentes dans les tissus humains l'est de manière stable. L'hydroxylation des 5mC ainsi que la régulation des protéines TET a été mise en évidence dans de nombreux mécanismes biologiques, notamment au cours du développement de l'embryon et du cerveau, ainsi que dans la pluripotence de certaines cellules. Des perturbations dans la régulation des protéines TET et dans leur activité d'hydroxyméthylation ont été identifiées dans deux grands types de pathologies : les cancers ainsi que les maladies neurodéveloppementales et neurodégénératives. Etant donné que la 5hmC et les protéines TET sont présentes dans la quasi-totalité des tissus cellulaires, il est probable que les recherches futures montreront leur implication dans de nombreux autres mécanismes biologiques et pathologies.

10. Projet de thèse

La découverte, en 2009, de la présence de 5hmC dans des quantités non négligeables chez les mammifères et que sa formation était catalysée par les enzymes TET à partir de 5mC, a immédiatement suscité l'intérêt de la communauté scientifique dans le domaine de l'épigénétique. Très rapidement, deux hypothèses ont été formulées quant au rôle biologique de cette réaction d'oxydation enzymatique. La 5hmC pouvait constituer un premier pas dans un processus de déméthylation active de l'ADN mais aussi être une modification épigénétique à part entière.

L'obtention d'éléments de réponses à ces deux questions fondamentales était limitée par les faibles quantités de 5hmC présente dans l'ADN, rendant sa détection difficile, et le fait que les principales techniques d'étude de la méthylation de l'ADN, déjà bien établies, ne pouvaient pas différencier la 5hmC de la 5mC. Afin de résoudre ce problème, les efforts de nombreux groupes de recherches se sont donc portés vers le développement de méthodes permettant la détection, l'enrichissement et la quantification de la 5hmC.

Lorsque j'ai commencé ma thèse, peu de méthodes pour l'étude de l'hydroxyméthylation de l'ADN avaient été publiées. L'équipe dans laquelle j'ai effectué ma thèse possédait une expertise sur l'étude de la méthylation de l'ADN et des inhibiteurs de DNMTs, et un projet d'étude de la 5hmC venait d'être initié. Des travaux préliminaires avaient ainsi été engagés pour une méthode de marquage enzymatique des 5hmC.

Dans ce contexte, mon projet de thèse a consisté à développer :

- (i) une méthode de quantification de la 5hmC dans l'ADN génomique en adaptant une méthode de quantification globale de la méthylation de l'ADN par spectrométrie de masse.
- (ii) des méthodes de marquage de la 5hmC en développant des approches chimiques et enzymatiques pour la fonctionnalisation spécifique de la 5hmC dans l'ADN.
- (iii) des outils pour identifier et caractériser des protéines partenaires interagissant avec la 5hmC, à l'aide de sondes oligonucléotidiques photoactivables.

Mon manuscrit s'articulera donc entre ces trois parties qui sont résumées dans la figure ci-dessous.

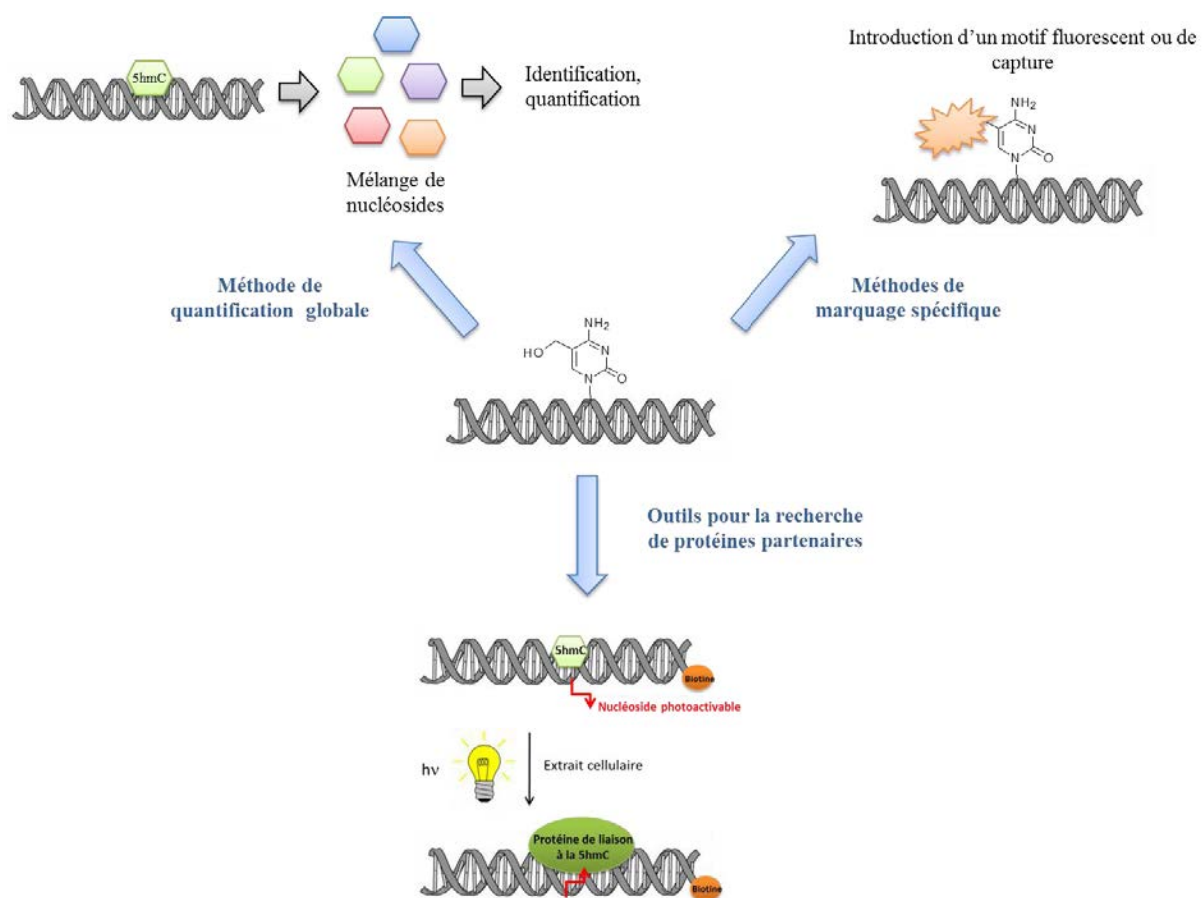


Figure 31. Différents axes du projet de thèse.

Chapitre II

Quantification globale par spectrométrie de masse : de la méthylation à l'hydroxyméthylation

En 2009, date de la (re)découverte de la 5hmC, peu de techniques étaient disponibles pour quantifier et localiser la 5hmC dans une séquence d'ADN. Comme nous l'avons vu dans l'introduction générale, l'identification de la 5hmC a été réalisée par des techniques classiques de séparation par chromatographie sur couche mince (CCM) de nucléotides radiomarqués au ^{32}P . Ainsi, la 5hmC a pu être initialement détectée et identifiée par une technique assez lourde qui consiste à (i) digérer l'ADN par une enzyme de restriction ; (ii) marquer les fragments obtenus au ^{32}P en présence de [α - ^{32}P] dGTP et d'une ADN polymérase ; (iii) digérer l'ADN par une nucléase ; (iv) analyser les nucléotides 3'-monophosphate radiomarqués par CCM en 2 dimensions. Cette technique, bien que sensible, est contraignante à mettre en œuvre.

Le développement d'outils simples permettant de quantifier de manière globale le taux de 5hmC dans des échantillons d'ADN est donc d'un grand intérêt. En effet, ces techniques pourraient permettre de comparer les taux de 5hmC dans différents types cellulaires ou dans des cellules à différents stades de développement, ou bien encore dans des cellules soumises à différents traitements, afin de mieux comprendre le rôle biologique de cette modification.

Au laboratoire, avant les publications de 2009 sur la 5hmC, la quantification globale de la méthylation de l'ADN génomique par spectrométrie de masse avait déjà été mise en place dans le cadre de travaux sur l'effet d'inhibiteurs de méthyltransférases de l'ADN dans un contexte cellulaire. Nous nous sommes donc penchés sur cette technique afin de voir si elle était aisément transposable à l'étude de la 5hmC dans des échantillons d'ADN extraits de cellules.

Avant de présenter ce que nous avons réalisé dans ce domaine, nous allons dans un premier temps présenter le principe général de la technique et de ses déclinaisons.

1 Quantification de la méthylation de l'ADN par spectrométrie de masse en tandem avec ionisation electrospray couplée à une chromatographie liquide

Les méthodes de quantification globale de la méthylation de l'ADN génomique sont basées sur des techniques de chromatographie telles que la chromatographie sur couche mince¹³⁸, la chromatographie en phase gazeuse¹³⁹, la chromatographie liquide¹⁴⁰, ou encore la chromatographie capillaire¹⁴¹. Elles reposent sur la séparation et la détection des composants de l'ADN (désoxyribonucléotides, désoxyribonucléosides, ou bases) obtenus par un traitement chimique ou enzymatique. Parmi ces techniques, les méthodes basées sur la chromatographie liquide sont les plus employées. Ainsi, le mélange peut être séparé sur phase inverse par HPLC et les quantités de 5mC quantifiées par spectrophotométrie UV mais cette méthode de détection comporte des points faibles liés à la sensibilité et à la nécessité d'avoir une excellente qualité de séparation.

Dans ce contexte, la détection par spectrométrie de masse a apporté des améliorations importantes permettant de résoudre ce problème de sensibilité. L'une des techniques développée dans ce domaine et reprise au laboratoire utilise la spectrométrie de masse en tandem avec ionisation electrospray couplée à une chromatographie liquide (LC-ESI-MS/MS) en mode MRM (Multiple Reaction Monitoring)¹⁴² afin d'analyser le mélange de nucléosides obtenu après digestion d'un échantillon d'ADN par un cocktail d'enzymes. Les enzymes utilisées pour la digestion sont : la nucléase P1, la phosphodiesterase de venin de serpent et la phosphatase alcaline. Ces enzymes coupent les liaisons phosphodiesters présentes entre deux nucléotides et hydrolysent les liaisons 5' et 3' phosphate conduisant à un mélange de nucléosides. Dans une expérience en mode MRM, des paires d'ions parent/fragment prédéfinis sont suivies spécifiquement et, bien que le mélange de composés analysés soit complexe, les spectres obtenus permettent de détecter uniquement les produits d'intérêts. Le mode MRM est ainsi particulièrement adapté à la quantification de petites molécules dans des mélanges biologiques et est très sensible.

Au laboratoire, cette technique a été reprise et adaptée sur la Plateforme de Protéomique et de Spectrométrie de Masse de l'UPMC en collaboration avec Gérard Bolbach

et Thierry Blasco. Cette adaptation a été réalisée sur un chromatographe Dionex-LC Packings Ultimate Plus couplé à un spectromètre de masse à triple filtre quadripolaire et piège ionique Applied Biosystem QTRAP à travers une interface de source d'ion TurboIonSpray¹⁴³. Elle a nécessité la caractérisation préalable de tous les nucléosides d'intérêt par spectrométrie de masse ESI afin de caractériser les ions parents et les fragments caractéristiques. La figure 32 présente les valeurs des masses des paires d'ions caractéristiques de chaque nucléoside sélectionnées en mode MRM pour la détection de chaque espèce. Elles correspondent aux masses $[M+H]^+$ de chaque nucléoside et de leur base azotée issue du clivage *N*-glycosidique.

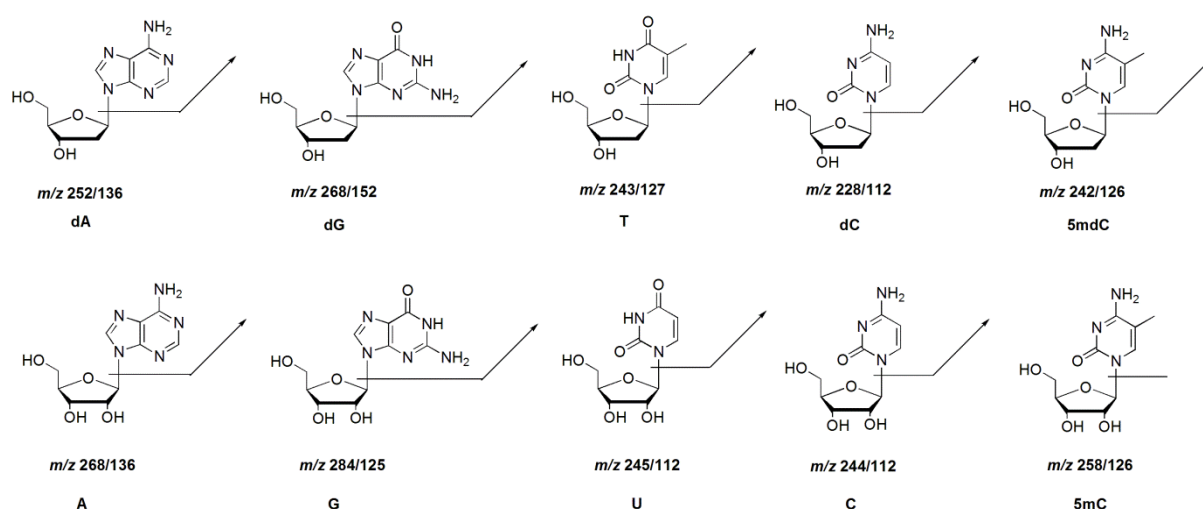


Figure 32. Masse des couples ions parents/fragments des différents nucléosides étudiés par LC-MS/MS MRM.

Le développement de cette méthode a également nécessité la mise au point de conditions de séparation par chromatographie liquide : celle-ci est effectuée sur une colonne Synergy 4u Polar-RP (Phenomenex) avec des solvants A (0.1% d'acide formique dans l'eau) et B (0.1% d'acide formique dans le méthanol) et un gradient de 0 à 50% du solvant B en 30 minutes. Les expériences préliminaires ont montré que la détection UV de la méthylation de la 5-méthyl-désoxycytidine est difficile pour des concentrations inférieures à la picomole. La détection par MS/MS en mode MRM permet un gain réel en limite de détection et en terme de quantification, soit par ajouts dosés d'un étalon, soit en choisissant l'un des nucléosides comme étalon interne. Pour la quantification, la désoxycytidine a été choisie comme étalon interne après avoir vérifié au moyen de courbes d'étalonnage qu'elle s'ionisait quasiment de la même façon que la 5-méthyl-désoxycytidine. La figure 33 montre à titre d'exemple les chromatogrammes respectifs obtenus pour un même échantillon d'ADN génomique extrait de

cellules HeLa par détection UV ou détection en MS/MS. On note que dans ces conditions, la 5mC n'est pas détectable en UV alors la spectrométrie de masse permet aisément de la détecter. Par ailleurs, la contamination d'un échantillon par de l'ARN (comme c'est le cas dans l'exemple présenté) ne constitue pas un obstacle car les nucléosides constitutifs se séparent bien dans les conditions utilisées.

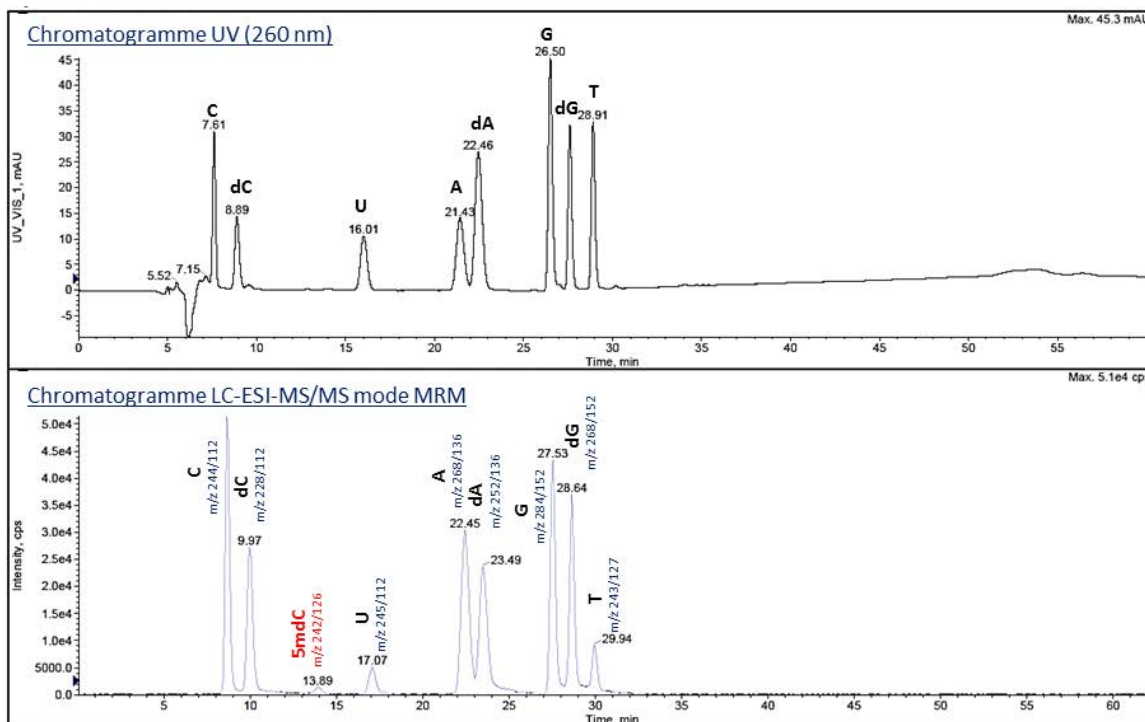


Figure 33. Chromatogrammes UV ou ESI-MS/MS d'un échantillon digéré d'ADN génomique issu de cellules HeLa.

Cette technique a été utilisée au laboratoire dans le cadre de différentes études sur les inhibiteurs de DNMTs afin de voir leur effet sur les taux de méthylation globaux dans un contexte cellulaire¹⁴⁴. Ayant en main ce savoir-faire, il m'a été proposé, dans le cadre d'une collaboration avec Joëlle Riond et Paola Arimondo (Unité mixte CNRS/Pierre-Fabre USR 3388 ETaC, Pharmacochimie de la régulation épigénétique du cancer à Toulouse), de participer à la validation d'une autre méthode d'analyse globale de la méthylation de l'ADN basée sur les anticorps. La technique par spectrométrie de masse étant une méthode directe, elle constitue une méthode de choix pour la validation d'une méthode indirecte.

2 Validation d'une technique de quantification de la méthylation globale d'échantillons d'ADN

Comme nous l'avons abordé dans l'introduction générale, la méthylation des cytosines est la modification épigénétique la plus étudiée chez les mammifères. Des perturbations des profils de méthylation sont observées dans certains types de cancers, avec une hypométhylation globale accompagnée d'une hyperméthylation au niveau de promoteurs de gènes suppresseurs de tumeurs. La méthylation de l'ADN étant un processus réversible, l'utilisation d'inhibiteurs des méthyltransférases de l'ADN pour permettre la réactivation des gènes suppresseurs de tumeurs constitue une des voies thérapeutiques contre le cancer. Si la capacité de certains inhibiteurs à diminuer le taux de méthylation *in vivo* est bien démontrée, le lien entre cette baisse de la méthylation globale et l'effet clinique anticancéreux n'est pas toujours bien compris. Le développement de méthodes permettant l'analyse fine des variations de taux de méthylation dans des cellules, à différentes étapes du cycle cellulaire, est donc d'un grand intérêt.

Dans ce contexte, nos collaborateurs à l'ETaC ont mis au point une méthode permettant l'étude simultanée de la méthylation et du cycle cellulaire dans des lignées cellulaires. Le principe est le suivant : les cellules à étudier sont perméabilisées puis incubées avec un anticorps primaire reconnaissant la 5mC, puis dans un deuxième temps avec un anticorps secondaire conjugué à un composé fluorescent. Enfin, les cellules sont marquées par de l'iodure de propidium, un agent fluorescent intercalant de l'ADN qui va permettre de déterminer l'étape du cycle cellulaire. Les cellules sont ensuite analysées par cytométrie de flux, une méthode qui permet la caractérisation individuelle de cellules en fonction de leurs paramètres physiques et de leur signal fluorescent. Il s'agit donc d'une méthode de quantification indirecte.

Afin d'évaluer cette méthode, Joëlle Riond s'est intéressée à l'effet d'un inhibiteur irréversible des DNMTs, la 5-azadésoxycytidine (5azadC), dont la structure est présentée dans la figure 34, sur le taux de méthylation de deux lignées cellulaires cancéreuses. Il s'agissait de cellules leucémiques (lignée KG1) et de cellules issues d'un mélanome (lignée WM-266). Ces

deux lignées cellulaires ont donc été traitées avec des concentrations croissantes de 5azadC pendant trois jours avant d'être préparées pour les analyses.

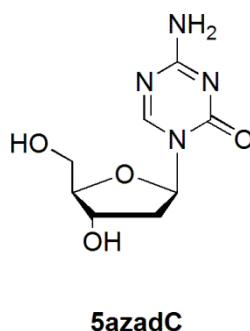


Figure 34. Structure d'un inhibiteur irréversible des DNMTs, la 5-azadésoxycytidine.

Pour valider cette nouvelle approche, nous avons effectué en parallèle l'analyse des taux de méthylation des échantillons d'ADN issus des cellules traitées par la méthode de spectrométrie de masse présentée précédemment, qui est une méthode de quantification directe. L'analyse par cytométrie de flux ne permettant d'accéder qu'à des valeurs relatives, nous avons normalisés les taux de méthylation que nous avons mesurés par rapport aux échantillons non traités. Les résultats des deux analyses sont résumés dans la figure 35.

Figure 35. Variation de la méthylation dans différentes lignées cellulaires lors d'un traitement par un inhibiteur de DNMTs. Les graphes A et B représentent les données issues de l'analyse par cytométrie de flux, pour les lignées WM-266 et KG1, respectivement. Les graphes C et D représentent les données issues de l'analyse par spectrométrie de masse.

On constate bien, par les deux méthodes d'analyse, que lorsque la concentration de 5azadC à laquelle les cellules sont traitées augmente, le niveau de méthylation diminue. Dans le cadre des cellules WM-266, les courbes obtenues sont similaires et montrent qu'aux plus hautes concentrations de 5-azadC, le taux de méthylation est réduit de moitié par rapport aux cellules non traitées. Notre analyse par spectrométrie de masse a donc bien permis de valider la méthode de cytométrie de flux développée par Joëlle Riond. Pour les cellules KG1, on constate cependant que, pour les concentrations les plus élevées de 5azadC, l'analyse par cytométrie de flux fait apparaître une diminution relative du taux de méthylation plus

importante que celle observée dans notre analyse par spectrométrie de masse. Cela peut s'expliquer par le fait que la cytométrie de flux est une méthode indirecte, qui fait entrer en jeu l'interaction entre les 5mC présentes dans l'ADN et l'anticorps qui va les reconnaître. La reconnaissance par cet anticorps peut être partiellement biaisée et dépendre de la densité locale en 5mC, ainsi que de l'accessibilité de celles-ci dans l'ADN. Comme ces données varient d'une lignée cellulaire à une autre, cela pourrait expliquer la différence observée.

Notre quantification par spectrométrie de masse a donc permis de valider la nouvelle méthode d'analyse simultanée de la méthylation et du cycle cellulaire développée par Joëlle Riond et Paola Arimondo. Ces résultats ont ainsi donné lieu à une publication récemment acceptée dans le journal *Epigenetics*¹⁴⁵.

3 La spectrométrie de masse pour l'étude de la 5hmC

Nous avons vu que la spectrométrie de masse est un outil puissant pour la quantification globale de la 5mC dans des échantillons d'ADN génomique. Qu'en est-il pour la quantification des bases oxydées dérivées de la 5mC ? Comme d'autres groupes, ayant déjà en main cette technique de détection, nous nous sommes tout naturellement intéressés à sa mise en œuvre pour la détection de la 5hmC.

3.1 Les études déjà réalisées

Très rapidement, la LC-MS a été mise en application pour l'étude de la 5hmC et l'un des premiers articles montrant l'apport de cette technique a été rapporté par le groupe de Thomas Carell dès 2010⁴⁷. Dans cette étude, la synthèse et l'utilisation d'isotopologues de la 5mdC et de la 5hmdC comme standard internes est rapportée pour une quantification précise de la 5hmC en combinant HPLC et une analyse par masse haute résolution (Figure 36). Nous avons déjà présenté dans le paragraphe 5.1 de l'introduction générale une application intéressante de cette technique qui a permis de quantifier les taux de 5hmC dans différentes zones du cerveau (Figure 17).

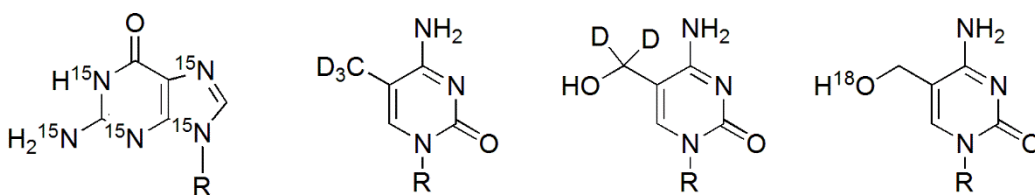


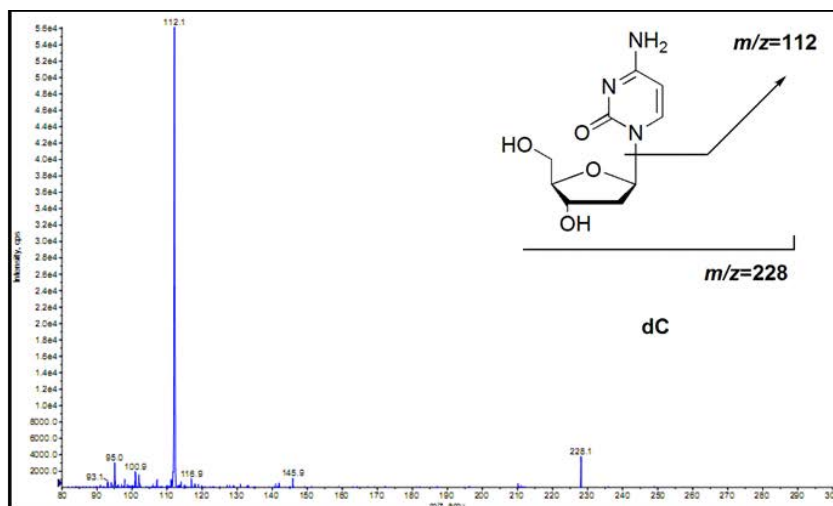
Figure 36. Différents isotopologues utilisées dans la quantification de la 5hmC et la 5mC par la spectrométrie de masse.

D'autres études ont également présenté l'utilisation d'isotopologues des nucléosides comme standards internes⁵² mais cette technique de quantification, si elle est très fiable comporte l'inconvénient majeur d'être très couteuse avec des composés peu accessibles. La 5fC⁵³ et la 5caC²⁵, présentes dans des quantités encore plus faibles que la 5hmC, ont aussi été détectées et quantifiées par la spectrométrie de masse.

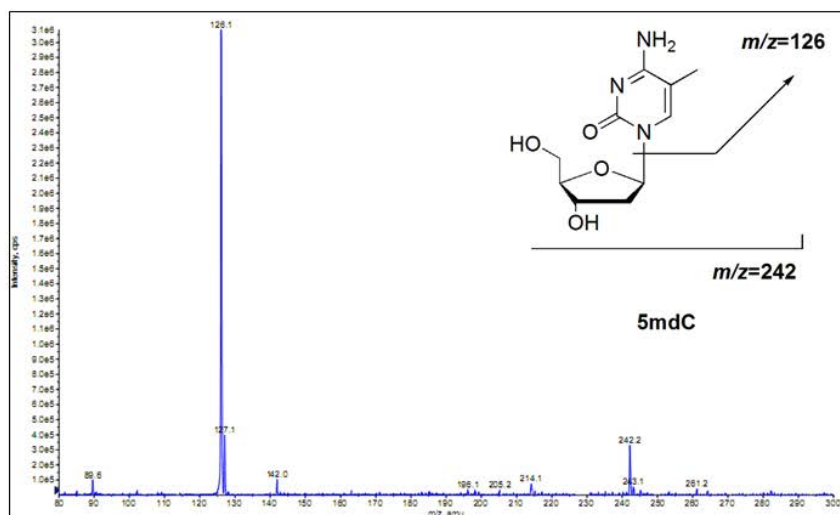
3.2 Mise au point de la technique au laboratoire

Dans un premier temps, les spectres de masse ESI des nucléosides d'intérêt ont été effectués afin de caractériser les ions parents et les fragments caractéristiques. Ce travail a été effectué à la Plateforme de Protéomique et de Spectrométrie de Masse de l'UPMC avec Thierry Blasco. Les nucléosides suivants ont été caractérisés : C, 5mC et 5hmC. Dans certaines études, la déoxyguanosine est utilisée comme standard interne lors de l'étude de duplexes d'ADN puisqu'en principe, dans l'ADN génomique $[dG] = [dC] + [5mdC] + [5hmdC]$, la présence de 5fdC et de 5cadC étant négligeable. Nous avons pour notre part choisi d'utiliser la déoxycytidine qui présente des caractéristiques d'ionisation et de temps de rétention plus proche de la 5hmdC et de la 5mdC. Les spectres et les paires d'ions caractéristiques de chaque composé sont présentés dans la figure 37.

A



B



C

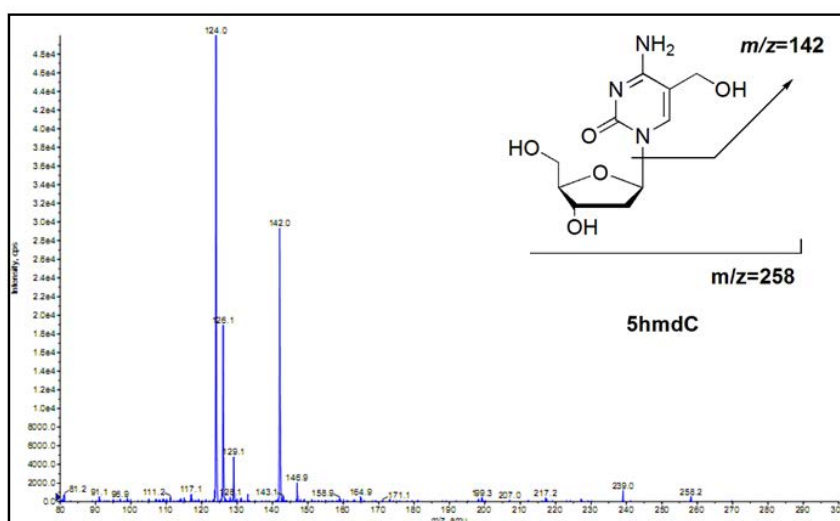


Figure 37. Spectres de masse et paire ions parents/fragments caractéristiques de **A)** la désoxycytidine ; **B)** la 5-méthyldeoxycytidine ; **C)** la 5-hydroxyméthyldeoxycytidine.

Dans un deuxième temps, un mélange de ces trois nucléosides a été analysé en LC-ESI-MS/MS en mode MRM permettant de suivre les trois paires d'ions associées en fonction de leur masse et de leur temps de rétention. La séparation est effectuée en phase inverse sur une colonne C-18. La Figure 38 présente un chromatogramme réalisé en LC-ESI-MS/MS en mode MRM. Comme cela avait déjà été reporté, la dC et la 5hmdC se séparent mal. Cependant, cela ne constitue pas un problème dans le cadre de la technique de détection par la spectrométrie de masse utilisée ici.

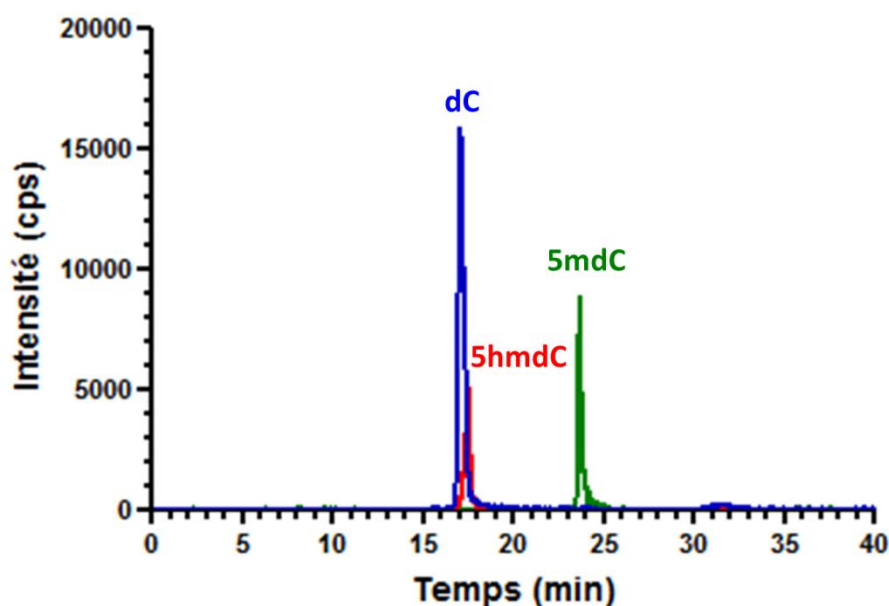


Figure 38. Chromatogramme LC-ESI-MS/MS d'un mélange des trois nucléosides dC, 5mdC et 5hmdC.

La reproductibilité et la limite de détection de la méthode a ensuite été évaluée en injectant des mélanges de trois nucléosides (dC, 5mdC, 5hmdC) à différentes concentrations. La figure 39 représente l'aire des pics associés aux différents nucléosides pour une gamme de concentration allant de 50 fmoles à 5 pmoles.

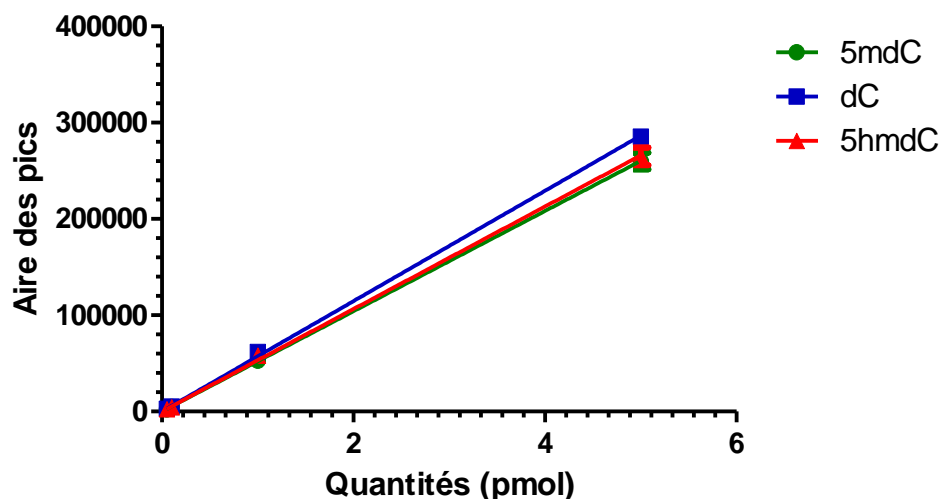


Figure 39. Gamme de concentration des nucléosides.

La répétabilité des essais est bonne et en ce qui concerne la sensibilité, la détection de 50 fmol de nucléoside est possible mais la détection de quantités inférieures est difficile.

Nous avons testé la méthode sur divers échantillons d'ADN génomique, extraits de différents types cellulaires fournis par plusieurs collaborateurs. La 5hmC ayant été détectée initialement dans de l'ADN extrait de neurones et de cellules souches embryonnaires de mammifère^{10,11}, nous avons choisi d'étudier de l'ADN de : (i) cellules neuronales de souris (fourni par le Dr Rajiv Joshi, CIRB, Collège de France), (ii) cellules embryonnaires de souris avant et après différenciation (fournies par le Dr. Pierre Antoine Defossez, Laboratoire Epigénétique et Destin Cellulaire, Paris VII). Nous avons également analysé différents échantillons d'ADN extrait de cellules humaines de cancers : DU145 (cancer de la prostate), HeLa (cancer du col utérin), WM-266 (mélanome) et KG1 (leucémie). Ces échantillons d'ADN nous ont été fournis par le Dr. Paola B. Arimondo à l'ETaC dans le cadre d'étude de l'ADN méthylé.

A chaque fois, 1 µg de ces échantillons d'ADN ont été traités par le protocole d'hydrolyse enzymatique décrit précédemment et les hydrolysats ont été analysés par la méthode de LC-ESI-MS/MS en mode MRM. Nous n'avons pas observé de 5hmC dans les cellules embryonnaires de souris, contrairement à ce que l'on pouvait attendre. Soit les échantillons que nous avons obtenus n'en contiennent pas, soit, la quantité de 5hmC est trop faible pour être détectée par cette méthode. Au contraire, la présence de 5hmC a été détectée dans l'ADN issus de cellules neuronales de souris, à un taux de 0.2% des cytosines totales, en accord avec les travaux de la littérature^{11,46}.

Les cellules cancéreuses présentent dans leur quasi-totalité des niveaux faibles de 5hmC, comme nous l'avons vu au paragraphe 5.1 de l'introduction générale. En accord avec ces données de la littérature, nous n'avons détecté qu'à de très rares occasions de la 5hmC dans les échantillons d'ADN issus des différentes lignées cancéreuses, dans des quantités très faibles (0.03% des cytosines totales). La figure 40 présente les chromatogrammes LC-ESI-MS/MS d'un tel échantillon d'ADN, issu de cellules WM-266.

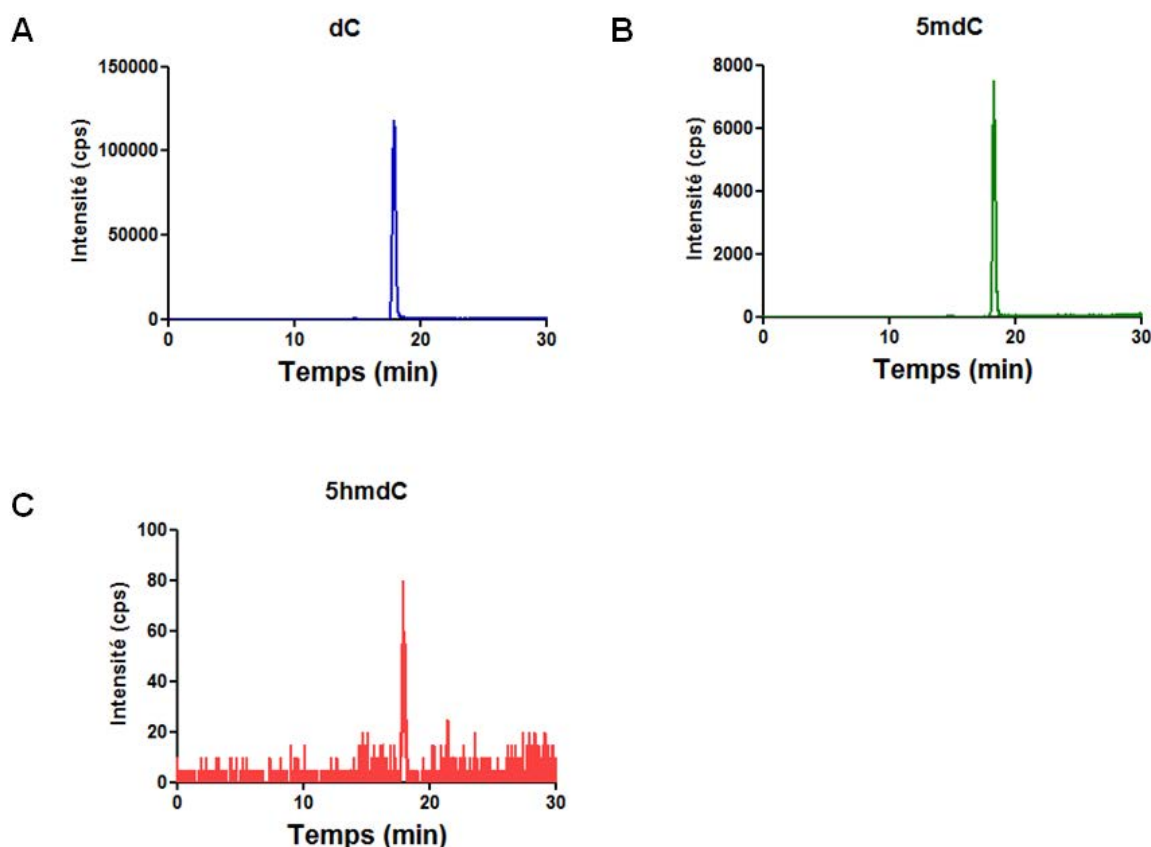


Figure 40. Détection de 5hmdC dans un des échantillons de cellules tumorales. Chromatogramme LC-ESI-MS/MS de la A) désoxycytidine B) 5-méthyl-désoxycytidine C) 5-hydroxyméthyl-désoxycytidine.

La méthode permet donc bien de détecter la 5hmC dans des échantillons d'ADN extrait de cellules mais étant donné qu'elle est présente en très faibles quantités, le signal lui correspondant est difficile à extraire du bruit de fond et rend problématique la quantification globale de la 5hmC.

En conclusion, la méthode d'analyse par LC-ESI-MS/MS en place au laboratoire pour l'étude de la méthylation de l'ADN ne semble pas adaptée, dans notre configuration, à la

quantification de la 5hmC. En effet, la faible présence de la 5hmC dans l'ADN nécessiterait de digérer de grandes quantités d'ADN pour dépasser les limites de détection, ce qui n'est pas forcément possible. Ce problème de sensibilité, couplé au fait que la digestion enzymatique des échantillons d'ADN ainsi que l'utilisation du spectromètre de masse représentent un coût non négligeable, en particulier pour une utilisation en routine, nous ont poussés à développer de nouvelles méthodes pour l'étude de la 5hmC, qui feront l'objet du chapitre suivant.

Chapitre III

Outils chimiques pour le marquage de la 5-hydroxyméthylcytosine dans une séquence d'ADN

A partir de 2009, il s'est rapidement posé la question de savoir comment différencier au sein d'une séquence les bases 5mC des 5hmC, puis, par la suite, le problème s'est étendu aux dérivés oxydés de la 5hmC. En effet, les principales techniques appliquées à la détection et au séquençage de l'ADN méthylé s'avèrent inappropriées pour discriminer les 5mC des 5hmC : ainsi, que ce soit le séquençage au bisulfite, l'utilisation d'enzymes de restriction ou encore les techniques d'immunoprécipitation ou d'immunofluorescence classiquement appliquées à l'ADN méthylé, aucune de ces techniques ne pouvaient être appliquées à l'étude de cette nouvelle base modifiée de l'ADN.

Afin de pouvoir répondre aux questions sur le rôle de la 5hmC dans l'organisme, en tant que modification épigénétique ou bien en tant qu'intermédiaire dans un processus de déméthylation de l'ADN, les efforts de nombreux groupes de recherches et sociétés se sont donc portés vers l'élaboration de méthodes pour la détection, l'enrichissement et le séquençage de la 5hmC dans l'ADN. Cette tâche était compliquée par le fait que la 5hmC est présente en très faible quantité dans l'ADN, ainsi que nous l'avons vu dans l'introduction générale.

L'un des objectifs de mes travaux de thèse débutés en 2011 a été de mettre en place de nouvelles stratégies de marquage de la 5hmC que je présenterai dans ce chapitre. Lorsque j'ai commencé cette étude, peu de travaux avaient été publiés, mais l'intérêt croissant qu'a porté la communauté scientifique dans le domaine de l'épigénétique à cette nouvelle découverte a conduit de nombreux grands groupes internationaux à travailler sur ce problème et une quantité importante de méthodologies ont été publiées tout au long de mon travail de thèse. Avant de présenter les travaux que j'ai réalisés dans ce domaine, je vais commencer par présenter de façon non exhaustive les différentes approches et améliorations technologiques qui ont été réalisées ces quatre dernières années pour quantifier et localiser la 5hmC ainsi que ses produits d'oxydation dans l'ADN.

1 Etat des lieux en 2014 des techniques développées pour l'étude de séquences d'ADN contenant la 5hmC

Toutes les technologies développées ne vont pas permettre d'accéder au même type d'information : certaines visent à quantifier globalement les taux de 5hmC dans différents échantillons, d'autres visent à faire de l'enrichissement de fragments d'ADN comportant ce motif, enfin d'autres techniques ont été développées afin d'accéder à la séquence précise de fragments d'ADN contenant la 5hmC. Ces techniques ont par ailleurs évolué et se sont sophistiquées pour ensuite permettre de discriminer entre 5hmC, 5fC et 5caC.

Les chapitres suivants seront donc consacrés à la présentation des différentes méthodes qui ont été développées depuis 2009 pour quantifier la 5hmC dans l'ADN ainsi que ses produits d'oxydations successifs identifiés ultérieurement. La plupart d'entre elles reposent sur l'utilisation de la réactivité spécifique de la 5hmC pour adapter des protocoles déjà en place pour étudier la 5mC.

1.1. Reconnaissance des séquences 5hmC par des anticorps

Une des techniques utilisées en biochimie pour la quantification de la méthylation de l'ADN est l'utilisation d'anticorps spécifiques de la 5mC. Dès 2010, les premiers anticorps anti-5hmC ont été commercialisés par la société Diagenode. D'autres ont été obtenus et commercialisés par la suite.

Ces anticorps peuvent être utilisés dans des applications d'immunofluorescence, où le couplage à un anticorps secondaire va permettre par exemple de visualiser la 5hmC dans des coupes de tissus et des cellules, ou bien de faire des quantifications relatives des niveaux de 5hmC dans des échantillons d'ADN génomique. Ils peuvent aussi être utilisés dans des expériences d'immunoprécipitation, et permettre ainsi d'identifier des séquences d'ADN contenant des bases 5hmC. L'anticorps est incubé avec de l'ADN génomique partiellement fragmenté et va se lier aux fragments contenant des 5hmC. Ces complexes anticorps-ADN sont ensuite récupérés par des billes paramagnétiques grâce à un anticorps secondaire approprié. L'étape d'élution finale permet de relâcher les fragments d'ADN qui pourra alors

être séquencé par une technique appropriée. On obtient alors des zones contenant une ou plusieurs 5hmC, sans pouvoir cependant en quantifier le nombre exact ou obtenir leur position précise.

Cependant, ces anticorps ne font pas l'unanimité car certaines études tendent à montrer qu'ils ne sont pas toujours spécifiques de la 5hmC et peuvent montrer certains biais de séquences. Ainsi, des faux positifs liés à une reconnaissance d'autres bases que la 5hmC peuvent survenir et par ailleurs, l'efficacité de ces anticorps semble dépendre fortement de la présence de régions denses en 5hmC. Enfin, il a été montré qu'il existait un enrichissement préférentiel pour les répétitions CA ou CT par les anticorps anti-5hmC, qui ne se retrouve pas dans d'autres méthodes d'enrichissement^{146,147}. Les anticorps anti-5hmC se révèlent donc peu efficaces dans les utilisations d'enrichissement¹⁴⁸. Par la suite de nouveaux anticorps reconnaissant l'adduit formé lors du traitement de la 5hmC par le bisulfite de sodium, la 5-méthylène sulfonatecytosine (CMS), ont été développés (Figure 41A). Ces anticorps anti-CMS^{124,148} dépendent beaucoup moins des séquences fortement enrichies en CMS (Figure 41B) puisque les auteurs de cette étude rapportent une reconnaissance possible avec en moyenne une 5hmC tous les 201 paires de bases¹²⁴.

Figure 41. A. Réaction du bisulfite sur la 5hmC conduisant au dérivé CMS ; **B.** Comparaison de l'efficacité des anticorps anti-5hmC et anti-CMS en immunofluorescence en fonction de la densité de 5hmC par fragment d'ADN¹²⁴.

Des anticorps ont aussi été développés pour reconnaître spécifiquement la 5fC et la 5caC et sont aujourd'hui accessibles commercialement.

1.2 Enrichissement de séquences par marquage chimio-enzymatique des 5hmC

A côté des techniques d'immunoprécipitation d'autres méthodes de type « pull-down » ont été développées pour permettre l'enrichissement de séquences. La quasi-totalité d'entre elles reposent sur l'utilisation d'une enzyme présentée dans l'introduction générale, la β -glucosyltransférase (BGT) du phage T4. Comme on l'a vu précédemment, la BGT permet de transférer un groupement glucose sur la 5hmC pour former la β -glucosyl-5-

hydroxymethylcytosine (glu-5hmC). Cette reconnaissance spécifique permettant d'introduire une nouvelle fonction sur les 5hmC *via*, par exemple, l'introduction d'un glucose modifié ou d'une modification du glucose après son introduction, a été exploitée, entre autres, dans la mise au point de trois méthodes permettant la capture et l'enrichissement de séquences d'ADN riches en 5hmC.

1.2.1 Hydroxymethylcytosine selective affinity labeling (hMe-Seal)

La première de ces méthodes, nommée Hydroxymethylcytosine selective affinity labeling (hMe-Seal), a été publiée par le groupe de Chuan He. Elle consiste à utiliser un analogue de l'UDP-Glucose dans lequel l'alcool primaire en position 6 du glucose est remplacé par un groupement azoture. La BGT est capable de reconnaître cet analogue et donc de transférer spécifiquement un motif glucose fonctionnalisable sur les 5hmC (Figure 42A)¹⁴⁹.

Figure 42. Enrichissement par la méthode hMe-Seal ; **A)** Transfert d'un motif glucose fonctionnalisable sur la 5hmC par la BGT ; **B)** Transfert du glucose modifié puis fonctionnalisation avec une biotine.

Après ce traitement enzymatique, les échantillons d'ADN sont fonctionnalisés par une biotine reliée à une dibenzylcyclooctyne par un bras espaceur (Figure 42B). Le couplage par « click chemistry » s'effectue de manière bioorthogonale, dans des conditions douces (notamment sans catalyse au cuivre). Les 5hmC sont alors sélectivement marquées par un groupement biotine, qui va permettre des étapes supplémentaires de capture, purification et enrichissement. Après séquençage des fragments purifiés, on obtient une cartographie des 5hmC présentes dans l'ADN. Cette méthode est maintenant disponible commercialement sous le nom d'Hydroxymethyl Collector Kit par la société Active Motif.

Le même groupe a par la suite proposé une méthode pour l'enrichissement de séquences comportant des 5fC (Formylcytosine selective affinity labeling, fC-Seal) se basant sur le principe de la méthode hMe-Seal (Figure 43)¹⁵⁰. La première étape est une glucosylation des 5hmC avec l'UDP-glucose et la BGT, suivie d'une réduction sélective des 5fC présentes dans l'ADN par NaBH₄. Les 5hmC générées par cette réduction vont pouvoir être marquées avec le substrat UDP-glucose-6-azido dans un deuxième temps. La réaction de

chimie « click » entre cet azoture et un alcyne reliée à une fonction biotine par un bras espaceur contenant un pont disulfure permet d'effectuer les étapes ultérieures d'enrichissement et de séquençage. La sélectivité vis-à-vis de la 5fC est assurée par la première étape de protection des 5hmC déjà présentes dans l'ADN *via* leur glucosylation.

Figure 43. Enrichissement de séquences contenant la 5fC par la méthode fC-Seal¹⁵⁰.

1.2.2 Interaction avec la protéine JBP1

La protéine JBP1 (J-binding protein 1) est une protéine identifiée chez certains trypanosomes, qui possède une très forte affinité pour la base J, une base rare chez ces parasites qui comporte un motif uridine glucosylé en position 5 très similaire à la glu-5hmC. Le groupe d'Arne Klungland⁵¹ a choisi de tirer parti de cette interaction et de la similarité entre la glucosylation de la 5hmU et celle de la 5hmC⁴³. La méthode repose donc sur deux étapes successives : premièrement, la glucosylation des 5hmC présentes dans l'échantillon d'ADN, puis la capture de celui-ci avec des billes magnétiques fonctionnalisées avec une JBP1 recombinante (Figure 44)^{151,152}. Cette méthode enrichit spécifiquement les séquences d'ADN contenant une 5hmC et permet leur analyse ultérieure pour déterminer leurs séquences.

Figure 44. Enrichissement de séquences contenant la 5hmC avec la protéine JBP1¹⁵².

Une étude comparative a été réalisée sur les méthodes hMe-Seal, JBP1 et l'immunoprécipitation par des anticorps anti-5hmC¹⁴⁶. Tant la capture par les anticorps que par la méthode hMe-Seal ont permis de reproduire des profils similaires de séquences contenant des 5hmC, au contraire de la méthode JBP1, qui semble donc plus spécifique de certaines séquences que les deux autres méthodes.

1.2.3 La méthode GLIB (Glucosylation, periodate oxydation, biotinylation)

La dernière des méthodes repose, de manière similaire, sur une première étape de glucosylation des 5hmC présentes dans l'ADN. Celui-ci est ensuite traité avec du périodate de sodium, qui permet de réaliser la coupure oxydante des liaisons carbone-carbone des diols vicinaux du glucose conduisant à la formation de deux fonctions aldéhydes (Figure 45). La fonctionnalisation de ces aldéhydes par une biotine *via* une fonction hydroxylamine reliée par un bras espaceur permet l'introduction de deux groupements biotines sur les 5hmC présentes dans l'ADN, permettant ensuite de réaliser les étapes d'enrichissement et de séquençage¹⁴⁸. Un des désavantages de l'oxydation au périodate de sodium est la dégradation possible de l'ADN générant un bruit de fond important.

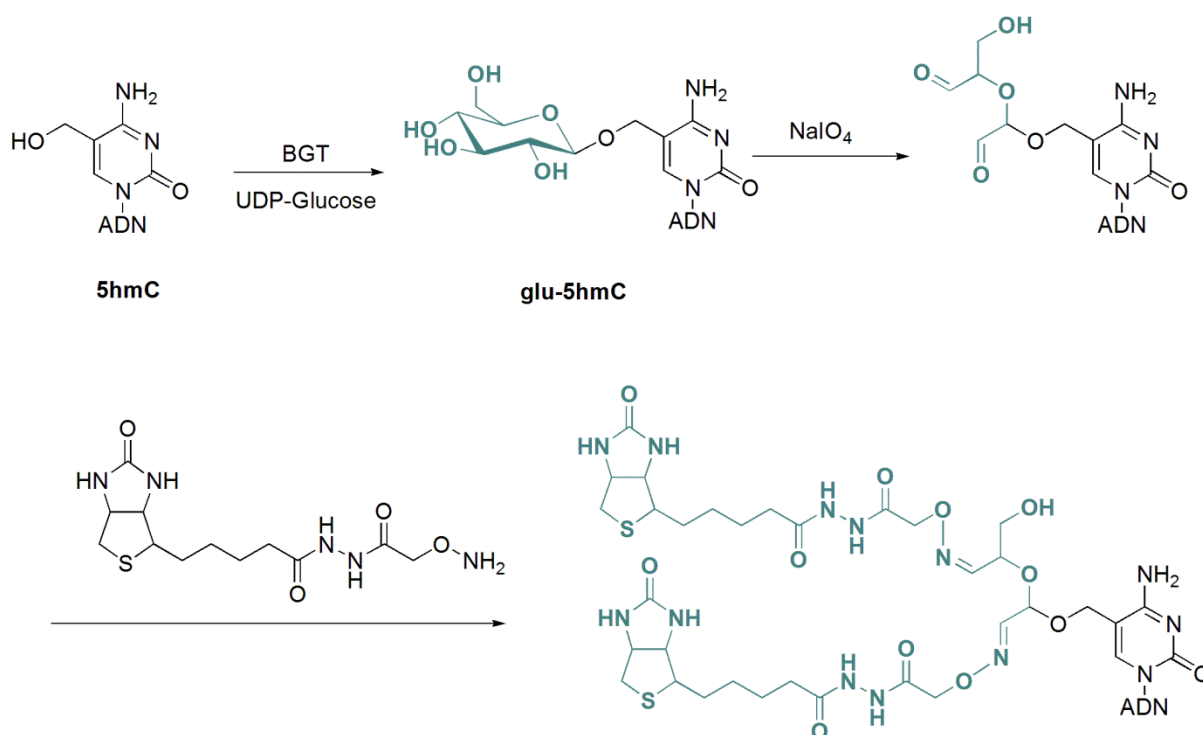


Figure 45. Principe de la méthode GLIB (glucosylation, periodate oxydation and biotinylation)¹⁴⁸.

1.3 Enrichissement de séquences par marquage chimique des 5hmC

La fonction alcool allylique permet également d'envisager de fonctionnaliser chimiquement les 5hmC présentes dans une séquence. Ainsi, le groupe de Hagan Bailey a rapporté récemment une méthode s'appuyant sur une réaction adaptée des méthodes de séquençage au bisulfite que nous évoquerons par la suite. Le traitement de la 5hmC dans des conditions douces par du bisulfite de sodium en présence d'un thiol permet ainsi la fonctionnalisation de la 5hmC¹⁵³. Le thiol utilisé est par exemple une biotine lié à une cystéine, ce qui permet comme précédemment l'enrichissement de séquences contenant des 5hmC (Figure 46). Cependant, le rendement de biotinylation est compris entre 35 et 55%, ce qui limite l'efficacité de cette méthode d'enrichissement qui n'a pas encore eu d'application concrète.

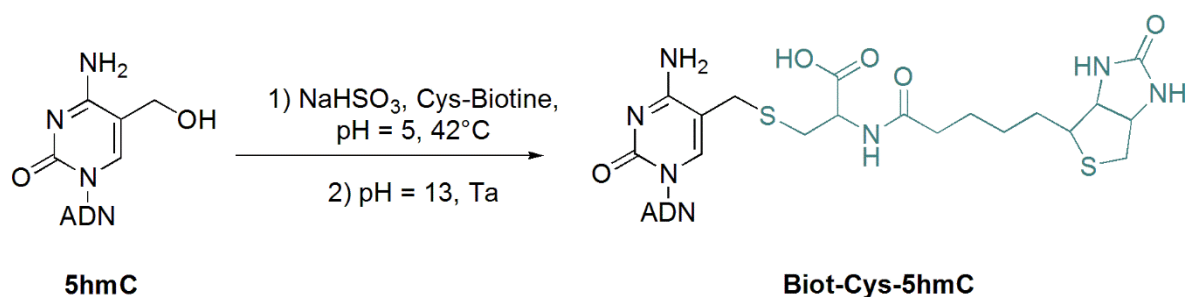


Figure 46. Fonctionnalisation de la 5hmC par un traitement au bisulfite en présence de cystéine-biotine¹⁵³.

Il est également possible de tirer parti de la fonction aldéhyde des 5fC pour introduire directement la biotine *via* un bras fonctionnalisé par une hydroxylamine afin d'enrichir des séquences contenant des 5fC (Figure 47)¹⁵⁴. Cette fonctionnalisation dans des conditions douces a permis un enrichissement 1000 fois supérieur aux résultats obtenus par immunoprécipitation des motifs 5fC et permis leur cartographie dans des cellules souches embryonnaires. Cependant, elle semble être moins spécifique des 5fC que la méthode fC-Seal

groupements méthyle et hydroxyméthyle. La recherche d'enzymes capables de reconnaître des séquences contenant des 5hmC a donc suscité un grand intérêt ces dernières années.

L'enzyme de restriction PvuRtsII peut ainsi cliver des séquences consensus contenant des 5hmC¹⁵⁵. L'enzyme a une préférence pour les séquences hydroxyméthylées sur les deux brins complémentaires du duplex d'ADN mais reconnaît aussi les duplex hémihydroxyméthylés. Cependant, cette reconnaissance des 5hmC semble dépendre d'une densité élevée en cytosines modifiées. D'autres enzymes présentant cette sélectivité de coupure pour des sites comportant la 5hmC ont été caractérisées, comme l'enzyme MspJI¹⁵⁶. Comme la digestion par ces enzymes produit des fragments d'une taille définie autour de la position modifiée, il est possible d'obtenir ensuite la position exacte de la 5hmC dans l'ADN.

Il est aussi possible de tirer parti du transfert spécifique d'un groupement glucose sur la 5hmC. L'introduction d'un groupement encombrant stériquement sur la 5hmC peut permettre de bloquer l'action d'enzymes de restrictions qui tolèrent la 5mC ou la 5hmC. Ainsi, l'enzyme MspI clive des séquences consensus (5' C[^]CGG 3') lorsque la cytosine interne est méthylée ou hydroxyméthylée mais n'agit pas lorsque cette même position contient une 5hmC glucosylée (glu-5hmC)¹⁵⁷. La différence de digestion par MspI entre des échantillons glucosylés peut permettre ensuite d'accéder à une cartographie de la position des 5hmC dans l'ADN (Figure 48).

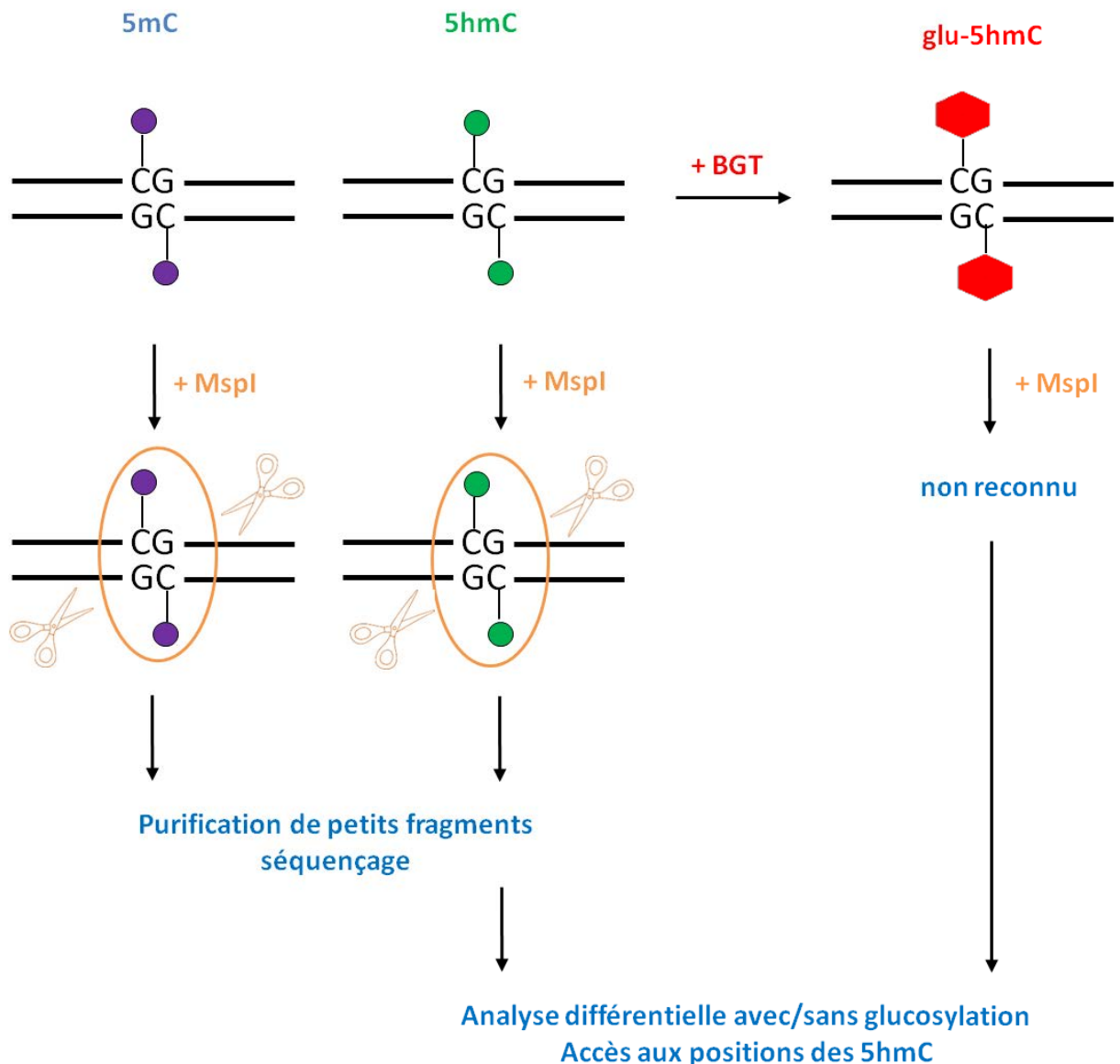


Figure 48. Principe de la cartographie des 5hmC à l'aide de l'enzyme de restriction MspI.

Une autre étude a montré que l'enzyme AbaSI, qui appartient à la même famille que PvuRtsII, clive préférentiellement les 5hmC glucosylées par rapport aux 5hmC non modifiées, et n'a pas d'activité sur les 5mC et les cytosines non modifiées¹⁵⁸. Cette enzyme a été utilisée dans une étude ultérieure pour cartographier les 5hmC dans des cellules souches embryonnaires de souris et valider son utilisation pour le séquençage¹⁵⁹.

En conclusion, l'utilisation d'enzymes de restriction permet d'obtenir des cartographies de la 5hmC dans des échantillons d'ADN en tirant parti des différences dans

leur tolérance vis-a-vis des cytosines non modifiées, des 5mC, et des 5hmC glucosylées ou non. Cependant, si les étapes de digestion de l'ADN par les enzymes de restriction sont simples et peu coûteuses, les étapes suivantes de purification des fragments digérés et de séquençage sont assez complexes à mettre en oeuvre. De plus, la portée du séquençage mis en oeuvre est limitée par le fait que les enzymes de restrictions sont spécifiques de séquences précises, et que seules les 5hmC présentes dans le contexte de ces séquences consensus seront mises en évidence.

1.5 Les techniques de séquençage

De nombreuses études ont été réalisées afin de mettre en place de nouvelles techniques de séquençage capables de distinguer la 5hmC et ses dérivés oxydés. En effet, jusqu'alors, le séquençage de l'ADN méthylé reposait sur la technique du traitement au bisulfite de sodium, qui, comme nous allons le voir s'est avérée inadaptée pour le séquençage de l'ADN hydroxyméthylé. Après un bref rappel sur la technique de séquençage au bisulfite, les principales méthodologies de séquençage mises au point jusqu'à aujourd'hui seront présentées.

1.5.1 Rappels sur le séquençage au bisulfite

Cette technique repose sur un traitement des séquences d'ADN avec du bisulfite de sodium qui forme un adduit à la position C6 de la cytosine (Figure 49A). Cet adduit est rapidement désaminé pour former l'uracile. Dans les cycles d'amplification qui suivent, cet uracile est alors remplacé par une thymine. Contrairement à la cytosine, la 5-méthylcytosine n'est pas sensible à l'action du bisulfite de sodium et elle sera remplacée par une cytosine lors des cycles d'amplification (Figure 49B). Après le traitement au bisulfite, les seules cytosines restantes correspondent donc aux 5-méthylcytosines initiales, ce qui permet, après analyse comparative, d'obtenir leur position précise dans la séquence d'ADN, ainsi que le taux de modification. Le séquençage au bisulfite est ainsi une méthode robuste et peu coûteuse, dont les principaux inconvénients viennent d'une conversion parfois incomplète de la cytosine en thymidine, ainsi que de la dégradation possible de l'ADN lors du traitement au bisulfite de sodium.

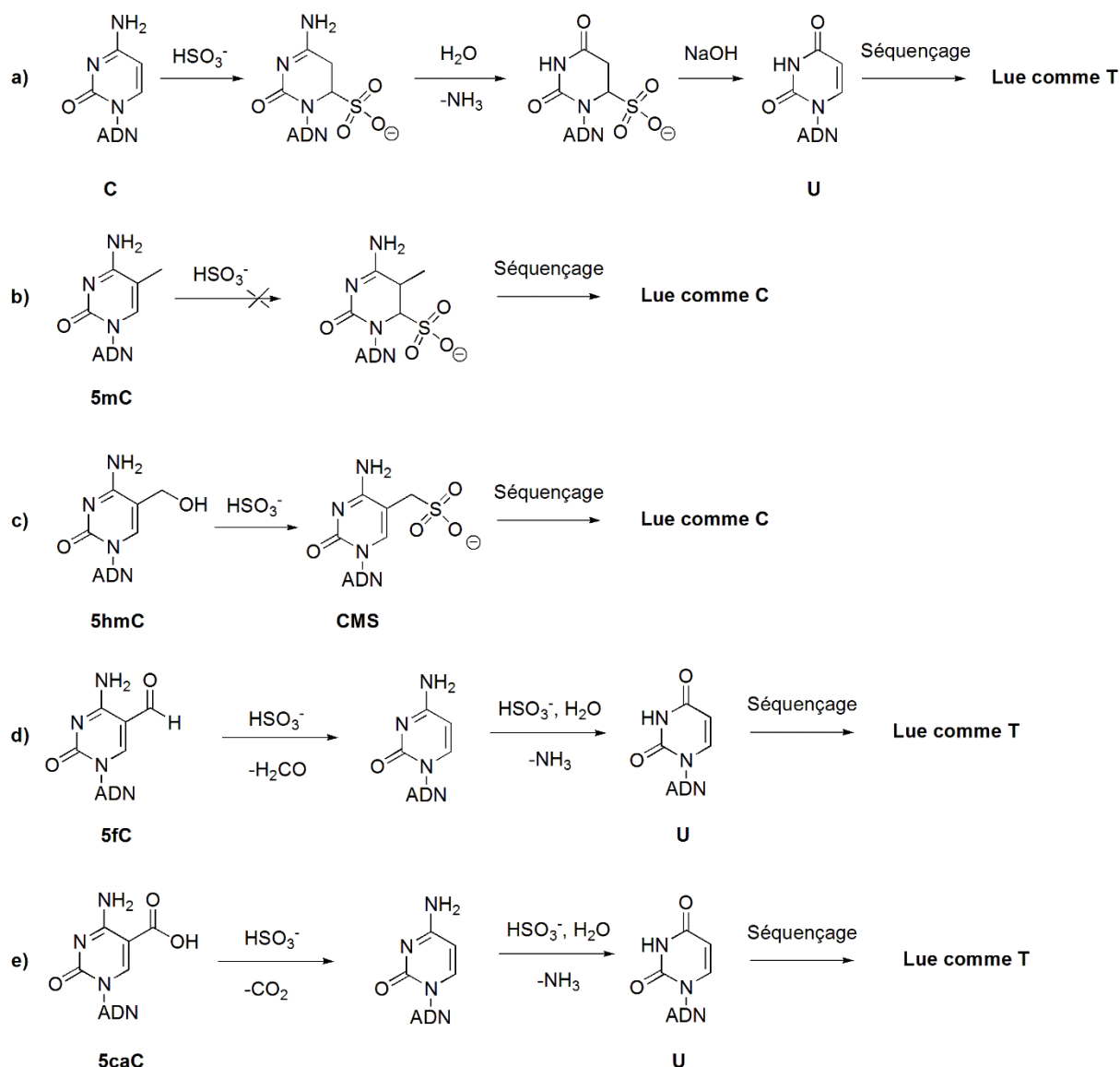


Figure 49. Les produits de réaction des nucléobases C, 5mC, 5hmC, 5caC et 5fC avec le bisulfite de sodium et l'interprétation lors du séquençage (d'après ¹⁶⁰).

Les découvertes successives des différents états oxydés de la cytosine ont mis à jour une autre faiblesse du séquençage au bisulfite : il ne permet pas de différencier entre 5mC et 5hmC d'une part, ainsi qu'entre 5fC, 5caC et C d'autre part^{161,162}. En effet, la 5hmC réagit avec le bisulfite conduisant à la formation d'un adduit stable (CMS) qui n'est pas désaminé. Elle sera lue comme une C lors du séquençage (Figure 49C). Au contraire la 5caC et la 5fC subissent respectivement une décarboxylation et une déformylation conduisant à la formation d'une cytosine qui est désaminée et lue comme une thymine lors du séquençage (Figure 49D et 49E). Afin de remédier à ce problème et de permettre le séquençage des différents états de

la cytosine par la méthode au bisulfite, de nombreux protocoles ont été développés qui seront présentés dans les paragraphes qui suivent.

1.5.2 Les techniques dérivées du séquençage au bisulfite

1.5.2.1 Les méthodes de séquençage OxBS et RedBS

Les deux méthodes dites OxBS-seq et RedBS-seq pour Oxidative Bisulfite Sequencing ou Reducing Bisulfite Sequencing, respectivement, ont été mises au point par le groupe de Balasubramanian et permettent les séquençages respectifs de la 5hmC et de la 5caC par la méthode au bisulfite. Leur principe est illustré dans la figure 50.

Dans la méthode OxBS-seq, la 5-hydroxyméthylcytosine est oxydée sélectivement en 5-formylcytosine par l'action du perrhuténate de potassium (Figure 50A). Comme nous l'avons vu, la 5fC est sensible au bisulfite. La comparaison des échantillons ayant subi le traitement au bisulfite normal, où la 5hmC va être lue comme une cytosine, et les échantillons ayant subi le traitement OxBS, où la 5hmC va être lue comme une thymidine, permet d'obtenir les positions de la 5hmC à la résolution de la base (Figure 50C). La conversion de la 5hmC en uridine (lue ensuite comme une T) dans le traitement OxBS a été estimée aux alentours de 95%, ce qui en fait une méthode fiable qui a été appliquée à l'étude de cellules embryonnaires, dont les résultats ont été présentés dans le paragraphe 5.3 de l'introduction générale⁶¹.

La méthode RedBS constitue une amélioration de la méthode précédente, publiée récemment⁶³, qui permet le séquençage des 5fC présentes dans des échantillons d'ADN génomique. En effet, le traitement par un réducteur, le borohydrate de sodium, permet de réduire les 5fC en 5hmC. La comparaison des échantillons ayant subi le traitement au bisulfite normal, où la 5fC va être convertie en U et lue en T, permet de déterminer la position des 5fC puisque celle-ci va être lue en C par la méthode RedBS (Figure 50B). Il faut noter cependant que, lors du traitement conventionnel au bisulfite, les auteurs ont estimé que 14% des 5fC étaient lues comme des C et pas comme des T. Cette conversion incomplète des 5fC aboutit à une surestimation du taux de 5fC dans les échantillons lors de l'analyse comparative avec les échantillons traités par la méthode RedBS.

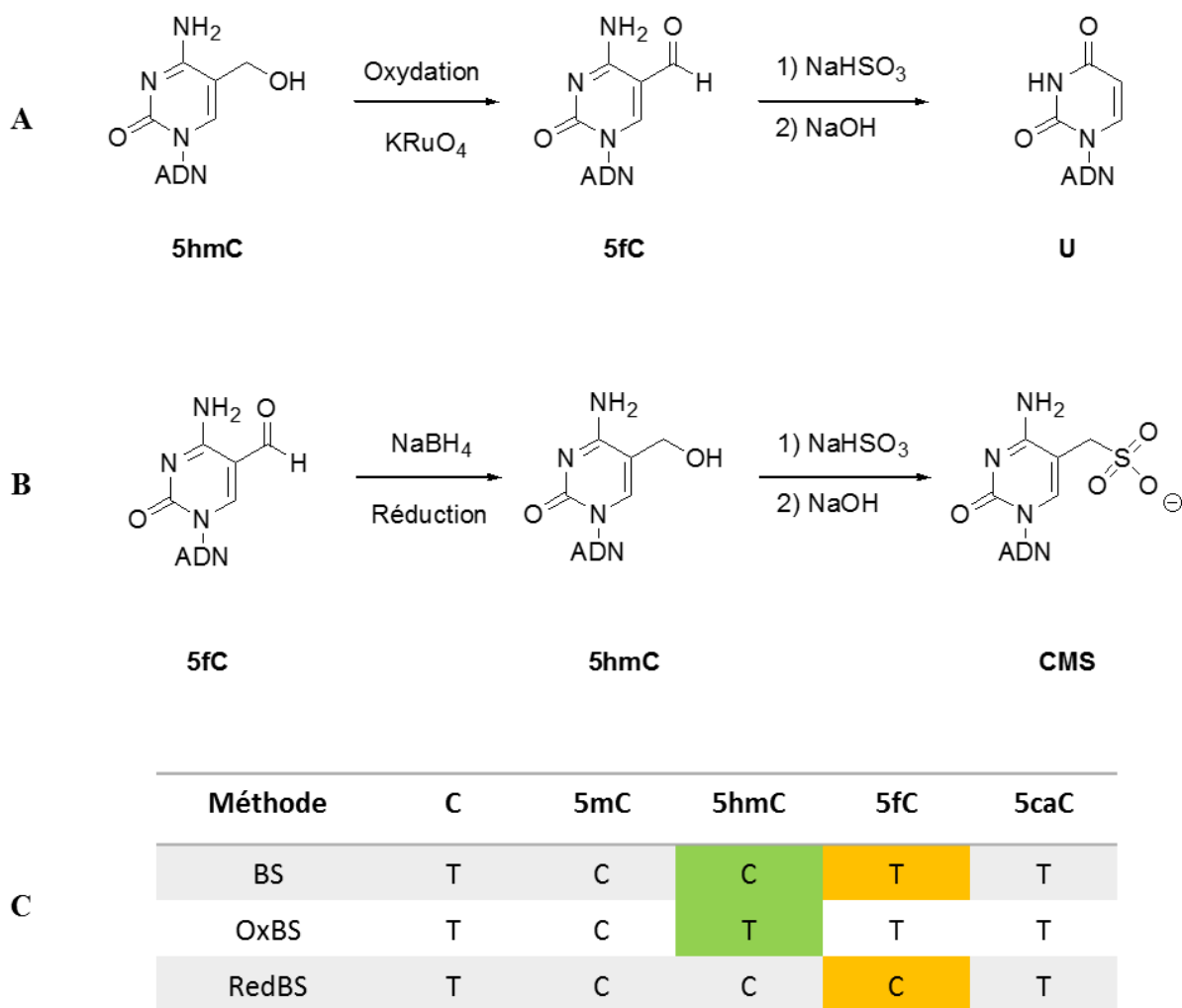


Figure 50. Adaptation du séquençage au bisulfite à la détection des 5hmC et des 5fC ; A) Principe de la méthode de séquençage OxBS ; B) Principe de la méthode de séquençage RedBS ; C) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.

1.5.2.2 La méthode de séquençage Tet-Assisted Bisulfite sequencing (TAB-seq)

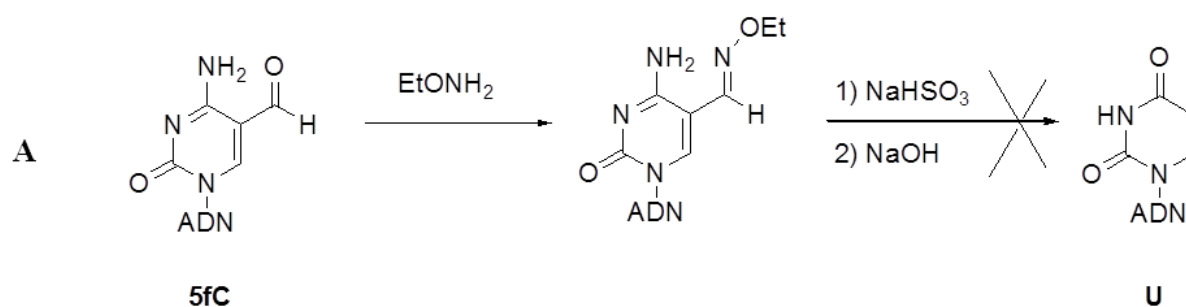
Une autre méthode pour séquencer les 5hmC consiste à utiliser la réactivité de la BGT, dont nous avons rappelé l'utilité dans le paragraphe 1.2 de ce chapitre, ainsi que les propriétés de l'enzyme TET1. Le principe de la méthode, nommée Tet-assisted Bisulfite Sequencing (TAB-seq), présenté dans la figure 51, est le suivant : un échantillon d'ADN génomique est d'abord traité par la BGT et l'UDP-glucose, ce qui permet de masquer toutes les 5hmC présentes en les glycosylant^{62,163}. L'échantillon est ensuite traité avec un excès du domaine

catalytique recombinant de l'enzyme TET1 de souris. L'enzyme TET1 oxyde les dérivés de cytosines non glycosylées par le traitement à la BGT, c'est-à-dire les 5mC et 5fC présentes dans les échantillons, conduisant à la formation des dérivés oxydés 5caC²⁵. La glu-5hmC étant la seule base qui n'est pas sensible au traitement au bisulfite (la C et les 5caC sont désaminées en présence de bisulfite), elle sera lue comme une C. Le taux de protection de la 5hmC par cette méthode a été évalué à 92% et constitue une limitation du TAB-seq, puisqu'il faut alors multiplier les séquençages pour attribuer les positions des 5hmC avec confiance.

Figure 51. Adaptation du séquençage au bisulfite à la détection des 5hmC ; A) Principe de la méthode de séquençage TAB-seq¹⁶³ ; B) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.

1.5.2.3 Les méthodes de séquençage fCAB-seq et CAB-seq

Le groupe de Chuan He a aussi développé une méthode pour séquencer la 5fC basée sur sa réactivité avec une hydroxylamine, nommée fCAB-seq¹⁵⁰. La réaction permet de protéger la 5fC de la conversion par la méthode du bisulfite. Encore une fois, la comparaison avec un traitement au bisulfite normal permet de trouver les positions de 5fC dans la séquence (Figure 52).

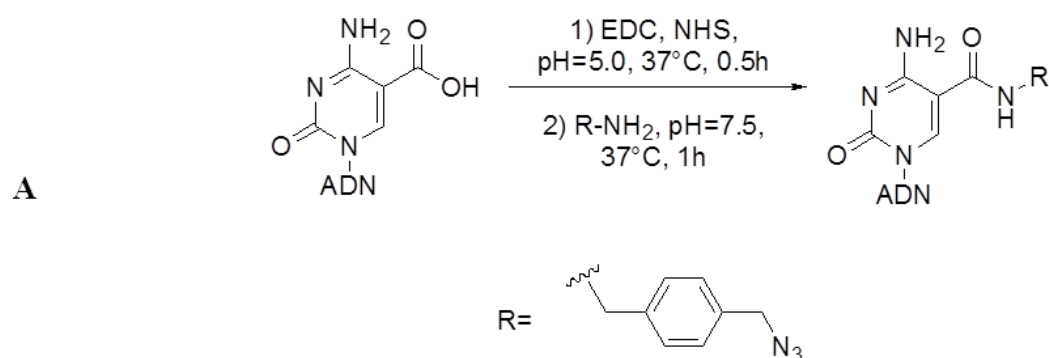


B

Méthode	C	5mC	5hmC	5fC	5caC
BS	T	C	C	T	T
fCAB-seq	T	C	C	C	T

Figure 52. Adaptation du séquençage au bisulfite à la détection des 5fC ; A) Réaction de la 5fC avec l'éthyle-hydroxylamine lors de la méthode fCAB-seq ; B) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.

Plus récemment, selon une approche similaire, une méthode de séquençage de la 5caC a été proposée par le même groupe, nommée CAB-seq¹⁶⁴. Cette méthode repose sur la réactivité spécifique de la 5caC avec un activateur d'acide carboxylique, le 1-Éthyl-3-(3-diméthylaminopropyl)carbodiimide (EDC). L'activation avec l'EDC suivie de la réaction avec une amine primaire permet la formation d'un lien amide. Les premiers essais de fonctionnalisation avec une amine primaire reliée à une biotine par un bras espaceur n'ayant pas donné de résultats, les auteurs ont utilisé l'azoture de 4-aminométhyle benzyle (Figure 53A). L'introduction de cette fonctionnalisation permet de bloquer la décarboxylation puis la désamination de la 5caC et de la rendre insensible au traitement par la méthode au bisulfite. La comparaison avec un traitement au bisulfite standard permet ensuite d'obtenir la position de la 5caC (Figure 53B).



B

Méthode	C	5mC	5hmC	5fC	5caC
BS	T	C	C	T	T
CAB-seq	T	C	C	T	C

Figure 53. Adaptation du séquençage au bisulfite à la détection des 5caC ; A) Couplage peptidique de la 5caC avec une amine, qui va permettre de la protéger lors du traitement au bisulfite ; B) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.

Une autre des applications de cette méthode est le pull-down de 5caC¹⁶⁴. Après activation avec l'EDC et fonctionnalisation avec l'azoture de 4-aminométhyle benzyle, il est possible d'introduire par une réaction de chimie « click » une fonction biotine qui va permettre l'enrichissement de séquences contenant des 5caC. Pour cela, les auteurs ont choisi une biotine reliée par un bras espaceur à une dibenzyl cyclooctyne (DBCO-biotine), un alcyne contraint qui permet de s'affranchir de la catalyse au cuivre (Figure 54)¹⁶⁵.

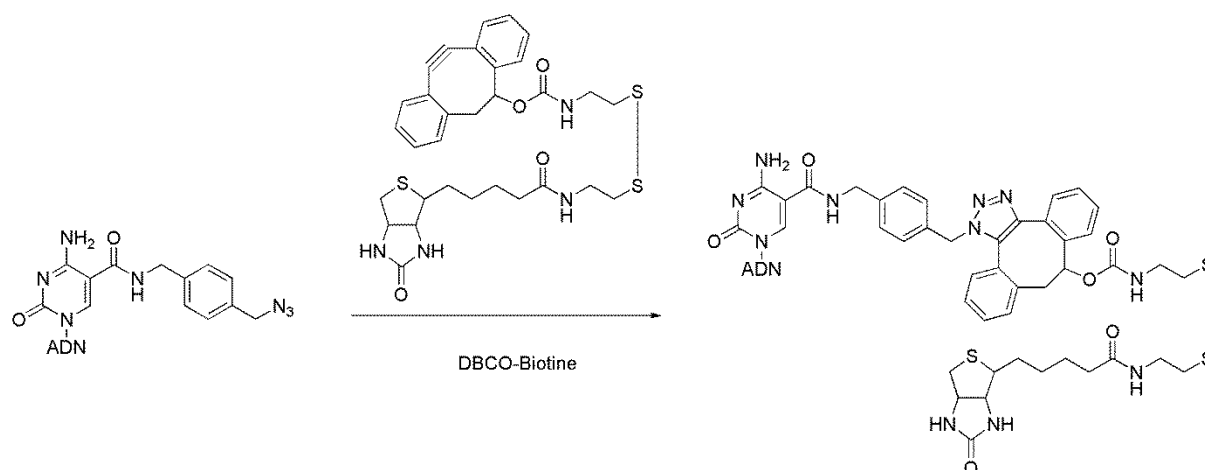


Figure 54. Fonctionnalisation de la 5caC par une biotine par la méthode CAB.

Ces deux méthodes n'ont pas encore été utilisées sur des échantillons d'ADN génomiques. La grande limitation est le faible taux de fonctionnalisation, au maximum de 65% avec l'amine présentée dans la Figure 53A, ce qui limite l'efficacité du pull-down ainsi que la fiabilité du séquençage de la 5caC, qui est comme on l'a vu précédemment présente dans l'ADN en quantités infimes. L'efficacité de la fonctionnalisation doit donc être améliorée avant d'utiliser cette méthode dans des applications réelles.

Des développements méthodologiques ont donc permis d'adapter le séquençage au bisulfite à la détection des différentes modifications de la cytosine. Cependant, ces méthodes ont des désavantages intrinsèques. Les différentes étapes de traitement et les taux de conversion ne sont pas toujours quantitatifs. Par ailleurs, les étapes d'amplifications successives peuvent induire des erreurs dans le séquençage.

Afin de remédier à ces problèmes, d'améliorer la vitesse et de diminuer le coût des séquençages, de nouvelles méthodes sont constamment en cours de développement. Parmi ces

méthodes, certaines reposent sur la lecture directe, non destructive, de fragments d'ADN non amplifiés, éliminant donc une partie des erreurs intrinsèques aux méthodes de séquençage précédentes. Parmi ces méthodes dites de « troisième génération », deux ont montré qu'elles pouvaient séquencer les bases modifiées de l'ADN et notamment la 5mC et la 5hmC : le séquençage par nanopores et le séquençage en temps réel de molécules uniques.

1.5.3 Le séquençage par nanopores

La méthode de séquençage par nanopores repose sur la détection de variations dans l'intensité du courant passant par un pore de taille nanométrique lorsqu'un brin d'ADN le traverse. Ces pores peuvent être de nature biologique (protéine membranaire dans une bicouche lipidique), de nature physique (pore déjà présent dans une membrane de silice) ou bien combiner les deux (protéines insérées dans un pore physique). Le dispositif expérimental met en jeu deux compartiments remplis d'électrolytes séparés par une membrane comprenant un pore, auxquels on va appliquer une différence de courant. L'application de ce voltage va entraîner le passage des fragments d'ADN préalablement déposés dans un des compartiments vers le deuxième compartiment, à travers le pore. Quand le brin d'ADN passe à travers le pore, il va bloquer plus ou moins le passage du courant en fonction de la taille des bases qui passent à travers le pore. La mesure de ce courant permet d'obtenir la nature de la base ainsi que, dans certaines applications, de séquencer le brin d'ADN passant à travers le pore.

Le groupe d'Hagan Bailey à Oxford, qui a contribué à développer cette méthodologie¹⁶⁶, a montré dès 2009 qu'il était possible par cette méthode de discriminer les différentes modifications des cytosines présentes sur des molécules d'ADN simples brins immobilisées dans des pores formés par l' α -hémolysine (Figure 55)¹⁶⁷.

Figure 55. Discrimination entre C, 5mC et 5hmC sur un simple brin d'ADN immobilisé dans un pore d' α -hémolysine¹⁶⁷. Un brin d'ADN (bleu) est immobilisé dans un pore d' α -hémolysine (grise) à l'aide d'un complexe entre une biotine (jaune) présente à l'extrémité 3' du brin d'ADN et une streptavidine (verte), qui va servir de « bouchon ». Trois sites de reconnaissance (R1, R2 et R3) peuvent permettre de distinguer entre les différentes bases non modifiées. L'introduction de cytosines modifiées dans la séquence d'ADN au site R1, mais pas au site R2, a permis d'observer des variations dans le courant traversant le pore suffisantes pour discriminer en C, 5mC et 5hmC.

Une autre équipe a montré qu'en utilisant des pores solides formés dans des membranes de nitrure de silicium, il était possible de discriminer des duplex comportant une base C, 5mC ou 5hmC toujours par des mesures de courant. Ce dispositif expérimental permet également de quantifier le pourcentage de 5hmC présent dans un duplex d'ADN¹⁶⁸.

L'équipe de Jens Gundlach a pour sa part, développé une autre méthode qui permet, elle, de séquencer des brins d'ADN en contrôlant leur passage à travers le pore d'une protéine bactérienne¹⁶⁹. En utilisant l'ADN polymérase du phage $\phi 29$, couplée à un mutant de la porine A de *Mycobacterium smegmatis* (MsPA), il est possible de contrôler le passage d'un simple brin de l'ADN à travers le pore de la MsPA à une vitesse suffisamment lente (vitesse médiane de 28 ms par nucléotide) pour permettre de résoudre les différents courants et de séquencer le brin d'ADN (Figure 56). Cette méthode a été appliquée au séquençage d'ADN comportant la 5mC ou de la 5hmC, permettant de les détecter à une résolution de l'ordre du nucléotide. Cependant, cette détection requiert l'utilisation d'algorithmes complexes ainsi que d'une comparaison avec une séquence non modifiée, qui peut être obtenue par PCR du fragment d'ADN original.

Figure 56. Séquençage d'un brin d'ADN passant à travers un nanopore¹⁷⁰. Un ADN simple brin (ssDNA) est entraîné à travers le pore de MspA (du compartiment trans au compartiment cis) par l'action de l'ADN polymérase $\phi 29$. Lorsque les bases passent au niveau du site de reconnaissance de MspA, à proximité du compartiment trans, on peut discriminer les différentes bases de l'ADN, y compris la 5mC et la 5hmC, en fonction des courants spécifiques associés.

Ces résultats sont très prometteurs et les développements technologiques très rapides dans le domaine. Cependant le séquençage par nanopore n'est pas encore accessible commercialement, que ce soit pour le séquençage des bases classiques ou celui des bases modifiées de l'ADN.

1.5.4 Le séquençage en temps réel de molécule unique (Single Molecule Real-Time sequencing, SMRT)

Cette méthode, développée par la société Pacific Bioscience, repose sur la détection à l'échelle de la molécule unique de sondes nucléotidiques fluorescentes lors de la polymérisation du brin d'ADN à séquencer. Elle a été nommée Single Molecule Real-Time

sequencing (SMRT). Le brin d'ADN à séquencer est lié à une ADN polymérase du phage $\phi 29$ immobilisée au sein d'une nanostructure appelée « zero-mode waveguide » (ZMW) qui va permettre la détection fluorescente de molécules uniques. Dans la solution sont présents des analogues des 4 désoxynucléosides triphosphate précurseurs de l'ADN portant chacun une sonde fluorescente émettant à une longueur d'onde différente sur le phosphate terminal. Quand l'ADN polymérase prend en charge puis incorpore un nucléotide dans l'ADN, le signal fluorescent est détecté à l'approche du nucléotide du site actif et disparaît lorsque le pyrophosphate fluorescent libéré diffuse hors du ZMW (Figure 57)¹⁷¹.

Figure 57. Principe de la méthode de séquençage SMRT. A) Le complexe polymérase-ADN fixé au sein d'un « zero-mode waveguide ». B) Incorporations successives des nucléotides et lecture des signaux correspondants¹⁷¹.

La longueur d'onde du signal fluorescent permet de déterminer la nature de la base incorporée (A, T, G, C), tandis que la durée de signal et l'intervalle avec le signal suivant permettent d'obtenir des renseignements sur une modification éventuelle de la base, en particulier la méthylation ou l'hydroxyméthylation de celle-ci. Ainsi, il a été montré que la polymérase met plus de temps à incorporer le nucléotide suivant lorsqu'elle va arriver à la position d'une 5mC, d'une 5hmC ou d'une 6-méthyladénosine (base méthylée présente chez les bactéries), et que les différences dans la durée des signaux ainsi que l'intervalle entre ceux-ci permettent de différencier ces bases modifiées de l'ADN¹⁷².

Cette méthode, qui nécessite tout de même de nombreux séquençages du même brin et la comparaison avec un échantillon non modifié, obtenu par PCR, a été améliorée par l'utilisation de la BGT¹⁷³. En effet, la fonctionnalisation des 5hmC présentes dans l'ADN par la BGT et l'UDP-6-azido-glucose permet de manière similaire aux techniques présentées précédemment dans le paragraphe 1.2 de venir greffer un bras espaceur relié à un groupement biotine par un pont disulfure sur les brins d'ADN. Ce groupement biotine permet l'enrichissement des séquences contenant une 5hmC. La réduction du pont disulfure permet de relarguer les fragments pour les séquencer par SMRT. La présence résiduelle d'un groupement encombrant stériquement sur la 5hmC permet d'augmenter grandement le signal et de se passer de la comparaison avec une séquence non modifiée.

Cette méthode permet l'incorporation (et donc le séquençage) de plusieurs milliers de paires de bases, à raison d'une à trois bases par seconde. Associé à la possibilité de combiner des dizaines de milliers de « zero-mode waveguide » par machine, cela en fait une technique très prometteuse pour le séquençage d'ADN contenant la 5hmC. Actuellement, l'inconvénient principal réside dans la présence de perturbations dans le signal qui sont observées non seulement à la position de la base modifiée mais aussi dans les bases suivantes (+2, 3, 6 pour la 5mC ; +2, 6 pour la 5hmC), ce qui complique l'analyse des résultats de séquençage et limite la résolution de celui-ci¹⁷².

Ces méthodes de séquençage dites de troisième génération permettent donc de détecter les bases 5hmC et des 5mC dans l'ADN. La méthode SMRT est déjà disponible commercialement, tandis que les nanopores nécessitent des améliorations supplémentaires avant de pouvoir être utilisés dans des applications de séquençages.

1.6 Les méthodes de quantification globale par marquage radioactif ou fluorescent

Nous avons pour l'instant vu des méthodes pour la détection, l'enrichissement ou le séquençage des 5hmC dans l'ADN. Un autre groupe de méthodes appliquées à l'analyse des bases modifiées de l'ADN est celui des méthodes de quantification(s) globale(s), qui vont permettre d'obtenir, à partir d'un échantillon d'ADN génomique, le taux de modification des cytosines (5mC, 5hmC, 5fC ou 5caC).

Une des premières méthodes proposées pour quantifier la présence de 5hmC dans des échantillons d'ADN a été de mesurer l'incorporation de glucose tritié en présence d'UDP-glucose-6-³H par la BGT¹⁷⁴. Après incorporation, précipitation de l'ADN et lavages successifs, il est possible de quantifier la radioactivité dans l'échantillon avec un compteur à scintillation. Cependant, la contrainte d'utiliser des substrats radioactifs n'a pas favorisé l'utilisation de cette méthode par la suite et les techniques de marquage fluorescent ont été privilégiées. Lorsque j'ai commencé à travailler sur ce projet, aucune technique n'était décrite dans ce domaine et cela a constitué l'un des objectifs de ma thèse. Je présenterai les études que nous avons réalisées dans le domaine et mentionnerai par la suite quelques études similaires qui ont été publiées au cours de ma thèse.

2 Développement de méthodes de marquage de la 5-hydroxyméthylcytosine

J'ai présenté dans les paragraphes précédents un grand nombre de méthodes développées depuis 2010 pour introduire sélectivement une modification sur l'alcool allylique de la 5-hydroxyméthylcytosine. Lorsque j'ai commencé mes travaux de thèse, le contexte était bien différent puisque la plupart de ces travaux n'étaient pas encore publiés.

L'une de ces approches qui venait tout juste d'être publiée, impliquant la BGT du phage T4 et l'UDP-6-azido-glucose, avait été aussi initiée au laboratoire et je me suis donc intéressé à sa mise en œuvre pour faire de la détection de la 5hmC par spectrométrie de fluorescence. Par ailleurs, je me suis également intéressé au développement d'une méthode de marquage chimique de la 5hmC, sachant qu'il n'en existait encore aucune lorsque j'ai commencé mon travail.

Le développement en parallèle de ces deux méthodes nous a paru pertinent car il pouvait permettre ainsi d'établir une étude comparative afin d'évaluer les rendements et la sélectivité d'un marquage enzymatique par rapport à un marquage chimique.

2.1 Marquage enzymatique des 5hmC

Dans l'introduction générale, nous avons présenté la β -glucosyltransférase (BGT) du phage T4. Cette enzyme catalyse l'introduction d'un groupement glucose sur l'alcool allylique en position 5 des 5hmC présente dans l'ADN du phage T4 (Figure 58).

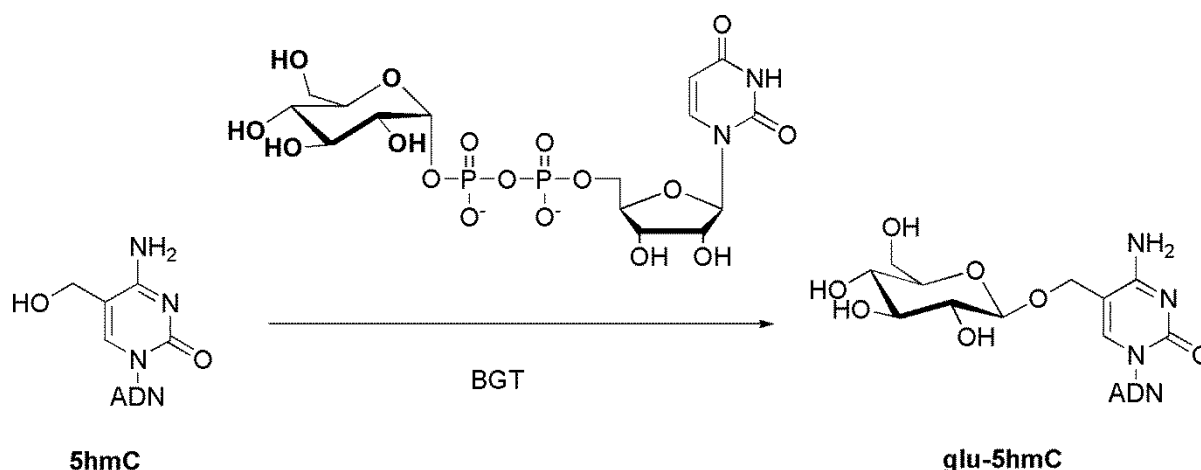


Figure 58. Glucosylation de la 5hmC par la BGT du phage T4.

Au laboratoire, une collaboration a été établie en 2010 avec le Dr. Solange Moréra (Laboratoire d'enzymologie biologique et structurale, CNRS, Gif sur Yvette), experte dans la structure des glucosyltransférases du phage T4¹⁷⁵⁻¹⁷⁸, afin de mettre en place une approche de marquage des motifs 5hmC à l'aide d'un analogue du substrat naturel de la BGT, l'UDP-Glucose, modifié chimiquement par une fonction azoture au niveau du glucose.

De nombreux groupes de recherches ont également pensé à appliquer la réactivité de cette enzyme dans des méthodes de quantification de la 5hmC et la compétition au niveau international a été très forte sur cette approche. Le premier exemple, présenté dans le rappel bibliographique précédent, a utilisé l'incorporation de glucose tritié par la BGT pour permettre une quantification des 5hmC¹⁷⁴. L'incorporation de glucose, modifié par une fonction bio-orthogonale destinée à la fonctionnalisation des 5hmC par un motif permettant un enrichissement ou une détection, a quant à elle été publiée pour la première fois par le groupe de Chuan He en 2011. Il a ainsi montré que la BGT était capable d'utiliser un analogue de l'UDP-glucose (Figure 59), l'UDP-6-N₃-glucose, pour transférer sur les 5hmC un glucose comportant un groupement azoture à la place de l'alcool en position 6¹⁴⁹. Par la chimie « click », il est alors possible de greffer les fonctionnalités désirées de manière bio-orthogonale. Dans le cas de cette étude, le greffage d'une biotine a permis de réaliser l'enrichissement de séquences d'ADN contenant des 5hmC en vue de leur séquençage.

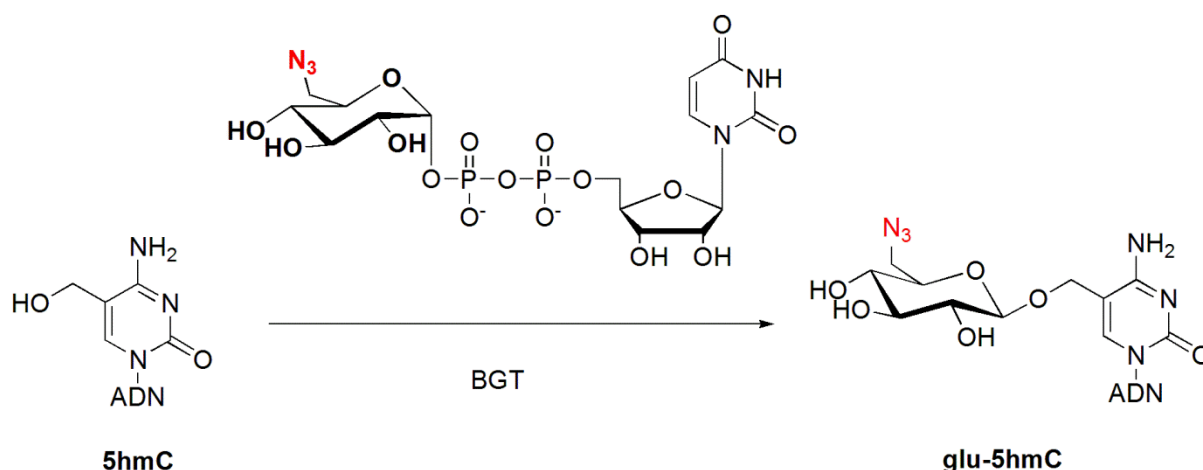


Figure 59. Fonctionnalisation des 5hmC par un sucre comportant une fonction azoture.

La synthèse de l'UDP-6-N₃-Glucose ayant déjà été amorcée au laboratoire avant la publication du groupe de Chuan He, je me suis donc intéressé au développement d'une nouvelle déclinaison de cette méthode permettant un couplage du groupement azoture avec un composé fluorescent afin de faire non pas de l'enrichissement de séquence mais de la détection et de la quantification de 5hmC par spectrométrie de fluorescence (Figure 60). Par ailleurs, afin de tenter d'apporter une plus-value par rapport aux précédents travaux, qui montraient que la BGT incorporait plus lentement le substrat modifié que son substrat naturel, un des objectifs annexes était de mettre en œuvre, en collaboration avec Solange Moréra, des techniques de cristallisation et de mutagenèse dirigées afin d'essayer d'améliorer le transfert du sucre modifié sur les 5hmC.

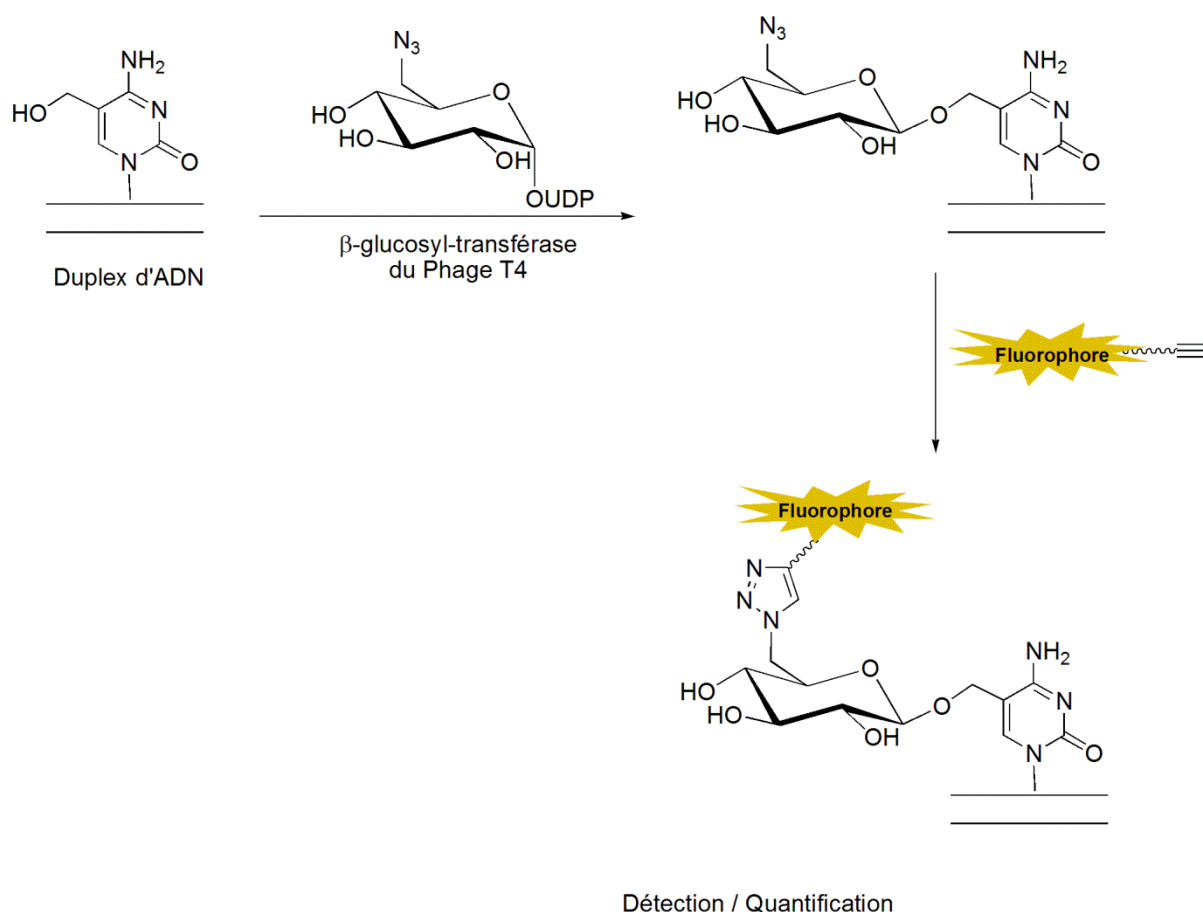


Figure 60. Méthode enzymatique de quantification de la 5hmC par fluorescence.

Pour mettre au point cette technique, nous avons donc dans un premier temps synthétisé ou produit les trois acteurs de cette approche : (i) un substrat oligonucléotidique contenant un motif 5hmC ; (ii) l'enzyme BGT et ses mutants ; (iii) le cofacteur donneur de glucose UDPG comportant un groupe azoture en position 6. Nous allons présenter dans les paragraphes qui suivent nos efforts dans l'obtention de ces différents outils.

2.1.1 Synthèse de la séquence oligonucléotidique

Pour développer cette technique de marquage, nous avons choisi de travailler sur une séquence synthétique d'ADN comportant la 5hmC. Nous avons opté pour un duplex d'ADN modèle de 30 paires de bases comprenant une 5hmC dans un contexte CpG positionné au milieu d'une séquence reconnue par les ADN méthyltransférases (Figure 61). Cette séquence nous a semblé être un modèle pertinent dans un contexte bibliographique où l'on ne

connaissait encore rien sur la localisation des 5hmC dans l'ADN. En effet, on savait à ce moment là que les 5hmC étaient issues de l'oxydation des 5mC qui sont elles-mêmes présentes dans les séquences reconnues par les DNMTs.

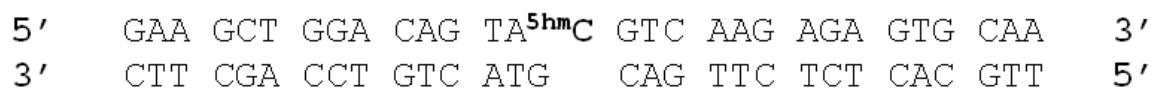


Figure 61. Séquence du substrat oligonucléotidique contenant une 5hmC.

La première séquence contenant la 5hmC sur laquelle j'ai travaillé avait été synthétisée sur support solide avant mon arrivée au laboratoire à l'Institut Pasteur sur la plateforme dirigée par Catherine Gouyette. J'ai par la suite réalisé moi-même la synthèse d'un certain nombre d'oligonucléotides à l'Université de Nice Sophia Antipolis, dans l'équipe du Dr. Rachid Benhida (Institut de Chimie de Nice, équipe Molécules Bioactives), dont celle de séquences hydroxyméthylées.

Le principe de la synthèse d'oligonucléotides sur support solide, présenté dans la figure 62, est le suivant : l'assemblage des nucléosides s'effectue à partir de l'extrémité 3' vers l'extrémité 5'. Le premier nucléotide est fixé sur un support de silice par une liaison covalente. Les nucléotides suivants sont généralement introduits par l'intermédiaire de synthons phosphoramidites protégés et fonctionnalisés de manière compatible avec le processus standard de synthèse d'oligonucléotides sur phase solide. Les bases peuvent par exemple comporter les groupements protecteurs suivants : phénoxyacétyle (Pac) pour les adénosines, *tert*-butylphénoxyacétyle (*t*BuPac) pour les guanosines et acétyle (Ac) pour les cytosines, et les phosphates sont protégés par un groupement cyanoéthyle (Figure 62). Dès la découverte de la 5hmC, la société Glen Research qui commercialise un grand nombre de synthons phosphoramidites a mis en vente un synthon 5hmC comportant pour groupement protecteur un groupement cyanoéthyle clivable en milieu basique en position 5 et un groupement benzoyl en position 4.

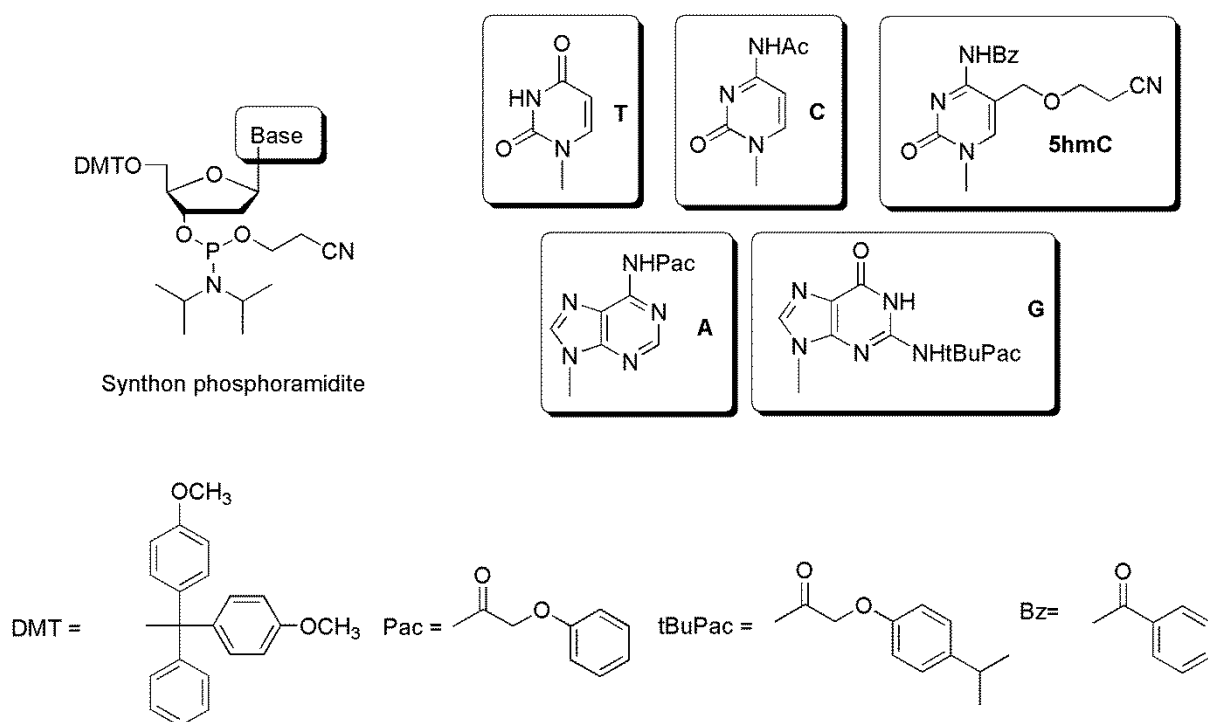


Figure 62. Structure d'un synthon phosphoramidite et des groupes protecteurs utilisés dans la synthèse.

A chaque cycle de couplage d'un nouveau synthon phosphoramidite, 4 étapes sont nécessaires (Figure 63) :

1) Détritylation

La première étape consiste à cliver le groupement protecteur diméthoxytrityle (DMT) en 5' du nucléoside fixé sur le support solide. Cette déprotection, effectuée par un traitement avec une solution d'acide trichloroacétique à 3% dans le dichlorométhane, conduit au dérivé OH-5' libre prêt à être couplé avec le synthon phosphoramidite lors de l'étape suivante.

2) Couplage

Au cours de cette étape, un intermédiaire tétrazolyl phosphoramidite activé est créé par addition simultanée du synthon phosphoramidite et du benzylthiotétrazole. Celui-ci sera à son tour substitué par l'hydroxyle-5' du nucléoside fixé sur le support solide pour former le lien phosphite internucléosidique.

3) Coiffage

L'étape de couplage n'étant pas quantitative, il reste toujours un faible pourcentage de nucléotides liés au support solide qui n'ont pas réagi. S'ils restent sous leur forme 5'-OH libre,

ils pourront subir l'étape suivante de couplage et engendrer ainsi une séquence différente de celle initialement prévue. Afin d'éviter ce problème, l'hydroxyle libre est acétylé en présence d'anhydride phénoxyacétique et de N-méthylimidazole. Cette étape minimise la présence d'impuretés et facilite ainsi la purification du produit final.

4) Oxydation

Lors de l'étape de couplage, le lien formé est de type phosphite triester (phosphore trivalent) instable et susceptible d'être clivé en milieu acide ou basique. C'est pourquoi, immédiatement après l'étape de coiffage, celui-ci est oxydé en phosphate triester pentavalent plus stable, en présence d'une solution aqueuse d'iode.

Le cycle est alors répété jusqu'à ce que l'élongation de la chaîne soit complète. Lorsque le dernier nucléotide a été introduit, l'oligomère formé est détaché du support de silice par un traitement alcalin (NH_4OH concentré, 12 heures, température ambiante), qui permet aussi d'éliminer les groupements protecteurs des bases. L'oligonucléotide peut alors être purifié par HPLC, sous forme de dérivé 5'-*O*-diméthoxytrityl (DMT-ON). Le groupement DMT est ensuite clivé par un traitement acide après purification par HPLC.

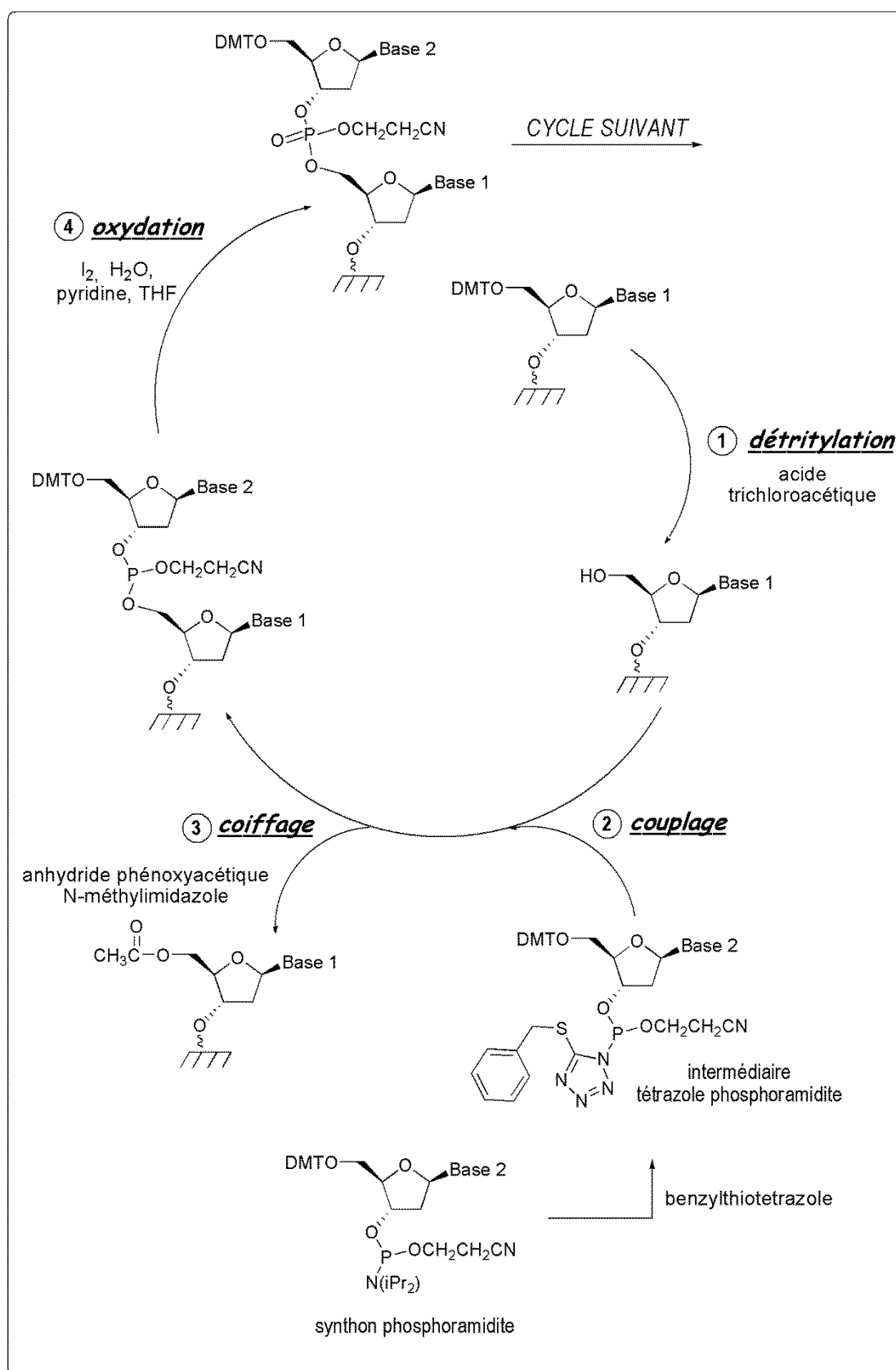


Figure 63. Différentes étapes de la synthèse automatique sur phase solide d'ADN.

L'oligonucléotide a été initialement caractérisé par spectrométrie de masse MALDI-TOF sur la plateforme de protéomique de l'UPMC indiquant que l'oligonucléotide avait une masse approximative de 9400 (masse calculée 9368), l'appareil utilisé ne permettant pas d'obtenir un spectre bien résolu.

Dans la suite de nos travaux avec ce premier oligonucléotide, nous avons rencontré quelques difficultés que nous mentionnerons par la suite. Quelques temps après, des travaux ont attiré l'attention sur le fait que le clivage quantitatif du groupement cyanoéthyle en position 5 des 5hmC nécessitait des conditions plus drastiques que celles indiquées par le fabricant. Une nouvelle analyse MALDI-TOF plus résolutive réalisée à l'Institut de Chimie des Substances Naturelles par Vincent Guérineau nous a permis effectivement de constater que l'alcool allylique de la 5hmC n'était pas entièrement déprotégé par le traitement classique en fin de synthèse de notre oligonucléotide (Figure 64). Nous avons donc traité l'oligonucléotide pendant 48h à 60°C, ce qui nous a permis de déprotéger entièrement cet alcool (Figure 63).

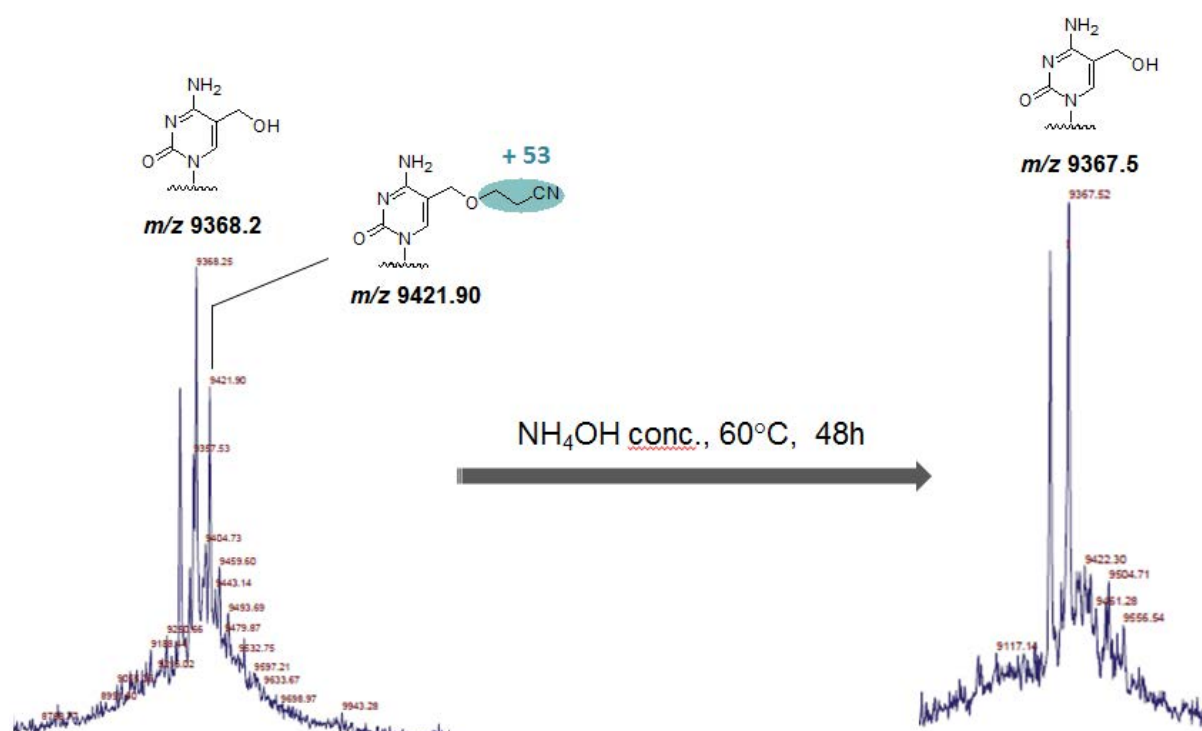


Figure 64. Déprotection totale de l'alcool allylique de la 5hmC.

La figure 65 présente le spectre de masse de l'oligonucléotide 5hmC avec lequel nous avons fait la plupart de nos études.

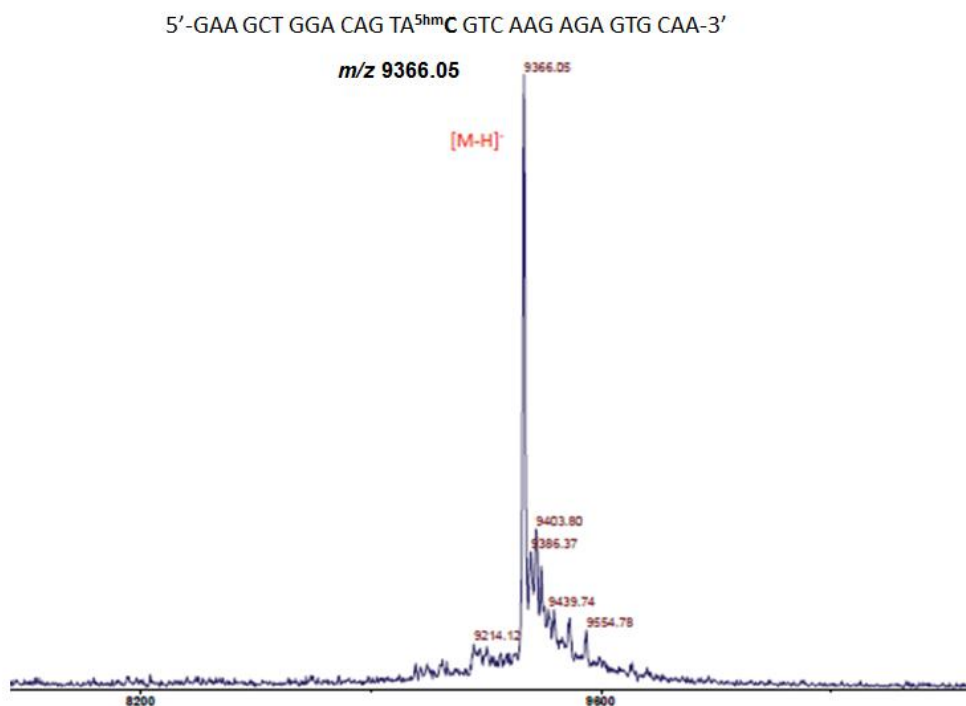


Figure 65. Spectre de masse de l'oligonucléotide contenant la 5hmC.

Depuis, d'autres synthons phosphoramidites pour la synthèse de la 5hmC ont été décrits et commercialisés. Notamment, la protection simultanée de l'alcool en position 5 et de l'amine en position 4 par un carbamate cyclique permet l'incorporation du phosphoramidite correspondant lors de la synthèse automatique, et une déprotection totale dans des conditions douces (NaOH 0.4M, 25 °C, 12 heures)¹⁷⁹.

De plus, certaines sociétés spécialisées en synthèse d'oligonucléotides ont proposé dans leur offre la synthèse d'oligonucléotides comportant la 5hmC. Nous avons d'ailleurs eu l'occasion d'en commander au cours de nos travaux.

2.1.2 Synthèse de la molécule UDP-6-azido-glucose

La synthèse du substrat UDP-glucose fonctionnalisé par un groupement azoture avait été initiée avant mon arrivée au laboratoire et avant que les travaux de Chuan He ne soient publiés¹⁴⁹ mais n'avait pas été finalisée. Il avait été choisi de fonctionnaliser la position 6 du glucose, ce qui en terme de synthèse constituait la voie la plus simple à mettre en œuvre, la question étant de savoir si cette fonctionnalisation n'aurait pas d'incidence majeure sur l'activité enzymatique. L'objectif était donc de synthétiser l'anomère α de l'UDP-6-azidoglucose même s'il a déjà été montré dans des réactions *in vitro* que la glucosyltransférase n'est pas inhibée par le mauvais anomère lors de l'utilisation d'un mélange des deux anomères α/β ¹⁷⁶.

Pour la synthèse des dérivés nucléoside 5'-diphosphate-sucre, deux stratégies de synthèse sont couramment employées (Figure 66)¹⁸⁰. L'une consiste à coupler un phosphate de glucosyle et un nucléoside 5'-monophosphate activé (stratégie A). Dans ce cas la stéréochimie au niveau du carbone anomérique est prédéterminée par la configuration de l'anomère du sucre monophosphate de départ. L'autre voie consiste à effectuer la glucosylation d'un nucléoside 5'-diphosphate (NDP) (stratégie B). Pour cela il faut synthétiser un sucre électrophile afin de pouvoir le coupler au NDP aisément accessible commercialement. Cette voie semble plus simple à réaliser mais présente l'inconvénient de donner lieu à la formation des deux anomères.

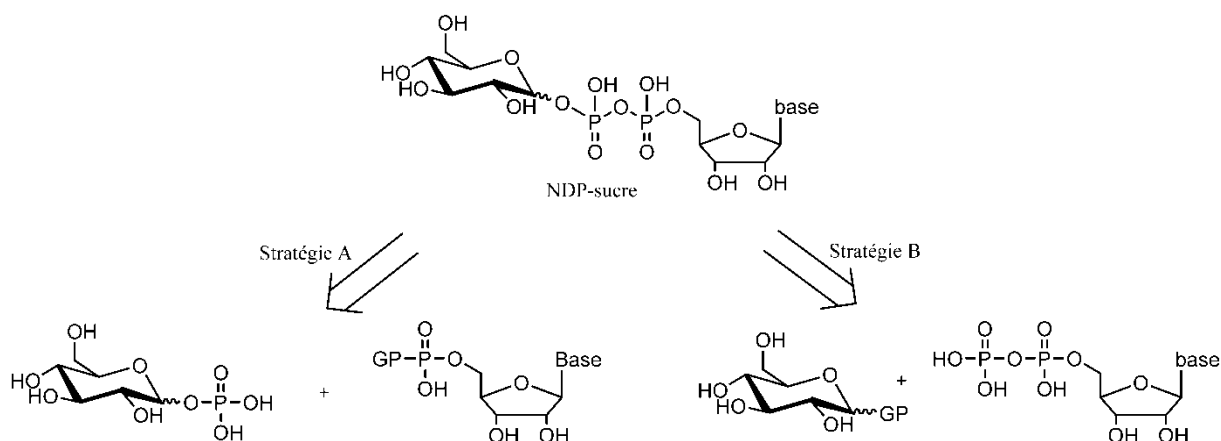


Figure 66. Stratégies générales de synthèse de NDP-sucre.

Nous avons opté pour la voie de synthèse A afin d'essayer d'obtenir l'anomère souhaité (α) majoritairement car nous souhaitons par la suite essayer d'obtenir des cristaux de la BGT complexée avec notre substrat modifié.

Nous avons donc besoin de deux synthons: un synthon azidoglucoyle phosphate que nous avons synthétisé et un synthon uridine 5'-monophosphate activée par un groupement morpholide en position 5' disponible dans le commerce. Nous allons décrire dans un premier temps la synthèse de l'azidoglucoyle phosphate, puis, le couplage entre les deux synthons.

• Synthèse de l'azidoglucoyle phosphate **7**

La synthèse de l'analogue de l'UDP-glucose est présentée dans les figures 67 et 68. L'alcool primaire en position 6 de l' α -D-méthylglucoside **1** est d'abord tosylé par l'action du chlorure de tosyl et de la triéthylamine dans la pyridine conduisant au dérivé tosylé **2** avec un rendement de 56%. Le groupement tosyl est alors substitué par l'action de l'azoture de sodium dans un mélange eau/acétone à 90°C pour former le composé **3** avec 75% de rendement. Ce composé **3** est ensuite acétylé aux positions 1, 2, 3 et 4 du sucre dans un mélange d'acide acétique et d'anhydride acétique avec de l'acide sulfurique concentré en quantité catalytique pour donner le composé **4**, avec un rendement de 75%. La déprotection sélective de la position anomérique de celui-ci est ensuite effectuée par la benzylamine dans le THF à 50°C, conduisant au composé **5** avec 90% de rendement. Pour introduire le groupement phosphate, nous avons choisi d'utiliser le 2-chloro-4H-benzodioxaphosphorin-4-one, qui donne lieu à la formation de l'anomère α^{181} . En présence de triéthylamine dans un mélange THF/1,4-dioxane, le 2-chloro-4H-benzodioxaphosphorin-4-one va permettre, après une étape finale d'hydrolyse, de former le composé H-phosphonate **6** avec 46% de rendement. L'ajout de tert-butyl hydroxyperoxyde et de diiode catalytique dans du THF permet d'oxyder le composé **6** et d'obtenir le phosphate **7** avec un rendement de 74%. La purification de ce dernier composé a été effectuée en phase inverse et il a été caractérisé par RMN notamment ^{31}P , confirmant la présence d'un seul anomère.

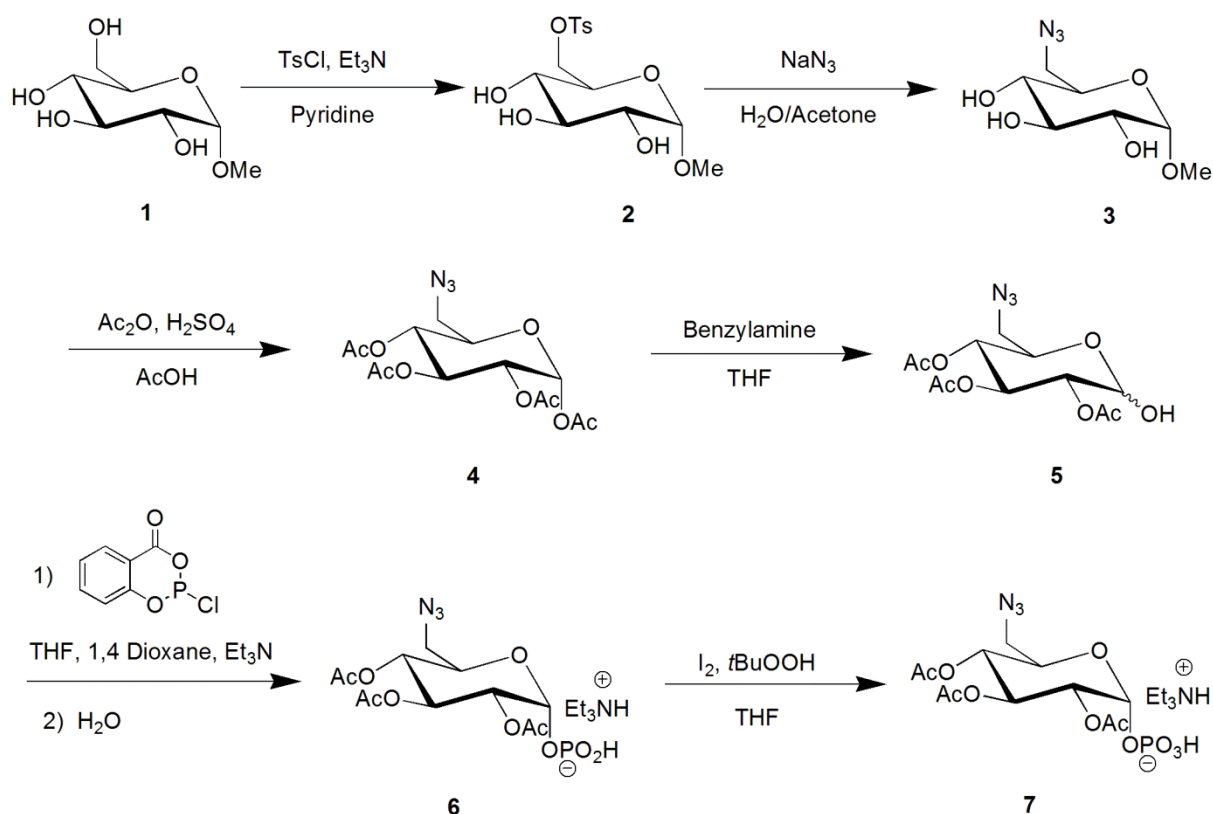


Figure 67. Synthèse du synthon azidoglycosyle phosphate **7**.

- Couplage du synthon azidoglycosyle phosphate **7** avec l'UMP-morpholidate

Une fois le phosphate **7** obtenu, nous avons réalisé le couplage avec l'UMP-morpholidate en présence de tétrazole dans la pyridine pour donner le composé **8** avec un rendement de 36%. Ce produit de couplage est enfin déprotégé en présence de bicarbonate d'ammonium et de triéthylamine dans le méthanol pour donner après purification par HPLC l'UDP-6-azido-glucose **9** avec un rendement de 85%. Cette synthèse en 8 étapes permet d'obtenir l'analogue désiré avec un rendement global de 3%.

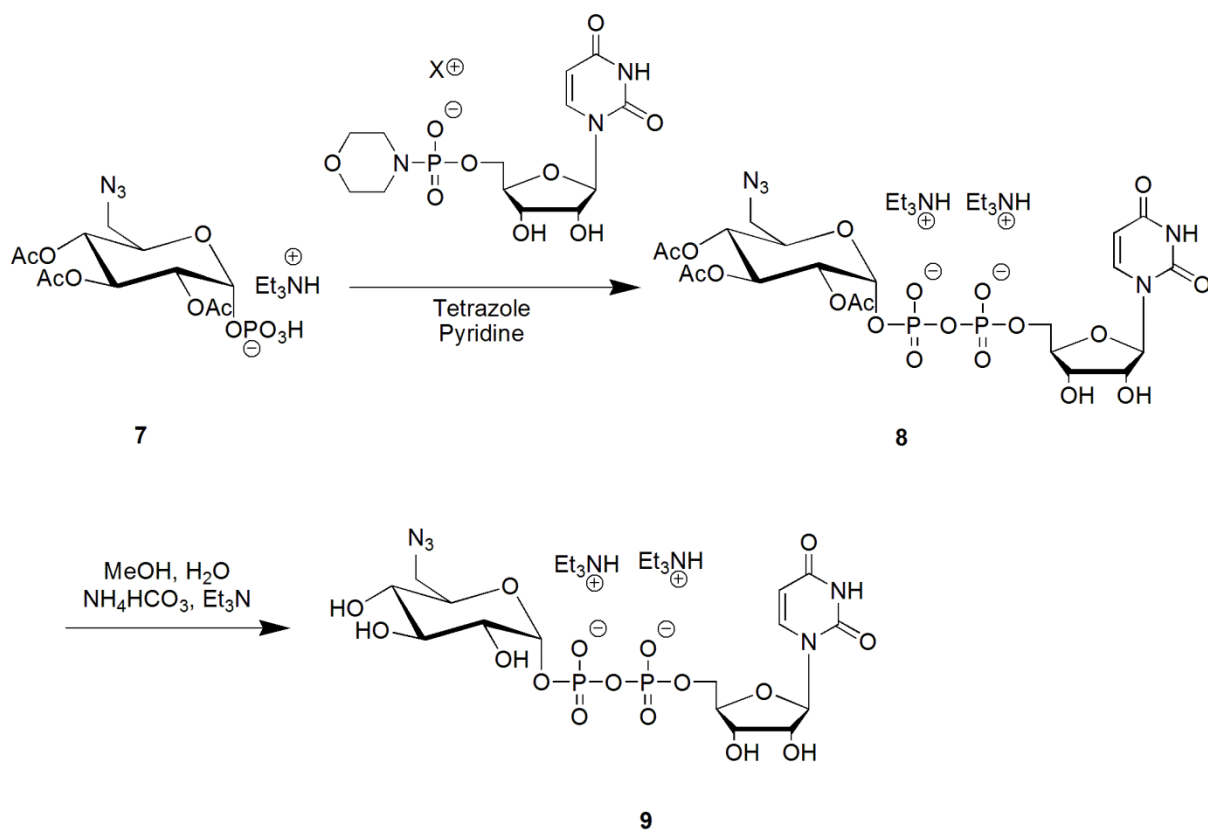


Figure 68. Synthèse de l'UDP-6-azido-glucose **9**.

2.1.3 Mise au point de la réaction de chimie « click »

Afin de mettre au point les conditions de fonctionnalisation du motif glucose par chimie click, nous avons travaillé sur l'un des intermédiaires dans la synthèse de l'UDP-6-azido-glucose, le composé **3**. Nous avons choisi comme fluorophore un composé disponible dans le commerce, l'Alexa Fluor 488 fonctionnalisé par un groupement dibenzylcyclooctyne (DBCO), l'utilisation d'un alcyne contraint permettant de réaliser la réaction de chimie « click » à faible température et sans qu'il soit nécessaire d'utiliser une catalyse au cuivre (Figure 69).

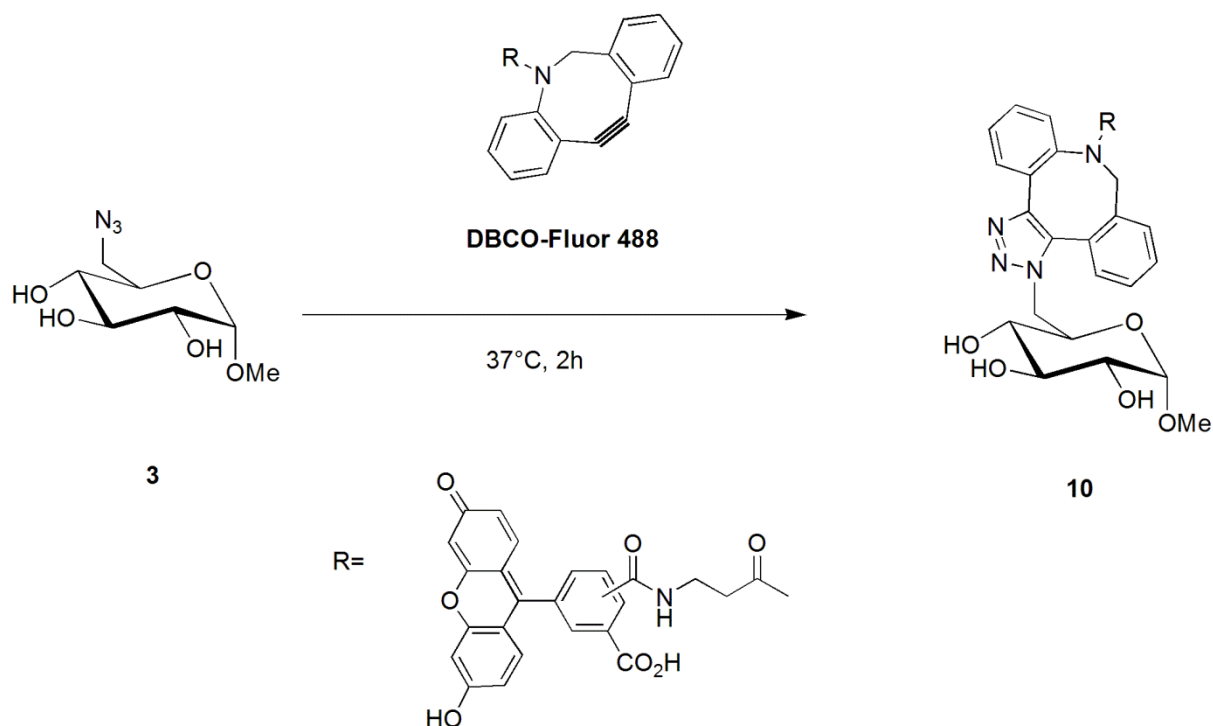


Figure 69. Fonctionnalisation du composé **3** par une réaction de chimie « click ».

Lors de l'analyse de la réaction par HPLC (Figure 70), nous avons observé que le composé DBCO-Alexa Fluor 488 commercial donnait lieu à 4 pics différents, qui ont été recueillis puis analysés par spectrométrie de masse MS/MS sur la Plateforme de Spectrométrie de Masse de l'Institut de Chimie Moléculaire de Paris Centre. La masse des 4 pics correspondait à la masse du DBCO-Alexa Fluor 488, nous avons donc émis l'hypothèse que le composé commercial était un mélange de régioisomères, ce qui n'était pas explicité par le fabricant (Jena Bioscience). Lorsque nous avons analysé le milieu réactionnel après deux heures de fonctionnalisation, nous avons observé la disparition des pics 3 et 4 et l'apparition de deux nouveaux pics, 5 et 6. L'analyse par spectrométrie de masse nous a permis de confirmer que les pics 5 et 6 correspondaient bien au composé **10**, tandis que les pics 1 et 2 étaient toujours composés du DBCO-Alexa Fluor 488. Certains des régioisomères ne semblent donc pas réactifs vis-à-vis du composé modèle **3**, ce qui nécessite d'utiliser le DBCO-Alexa Fluor 488 en large excès.

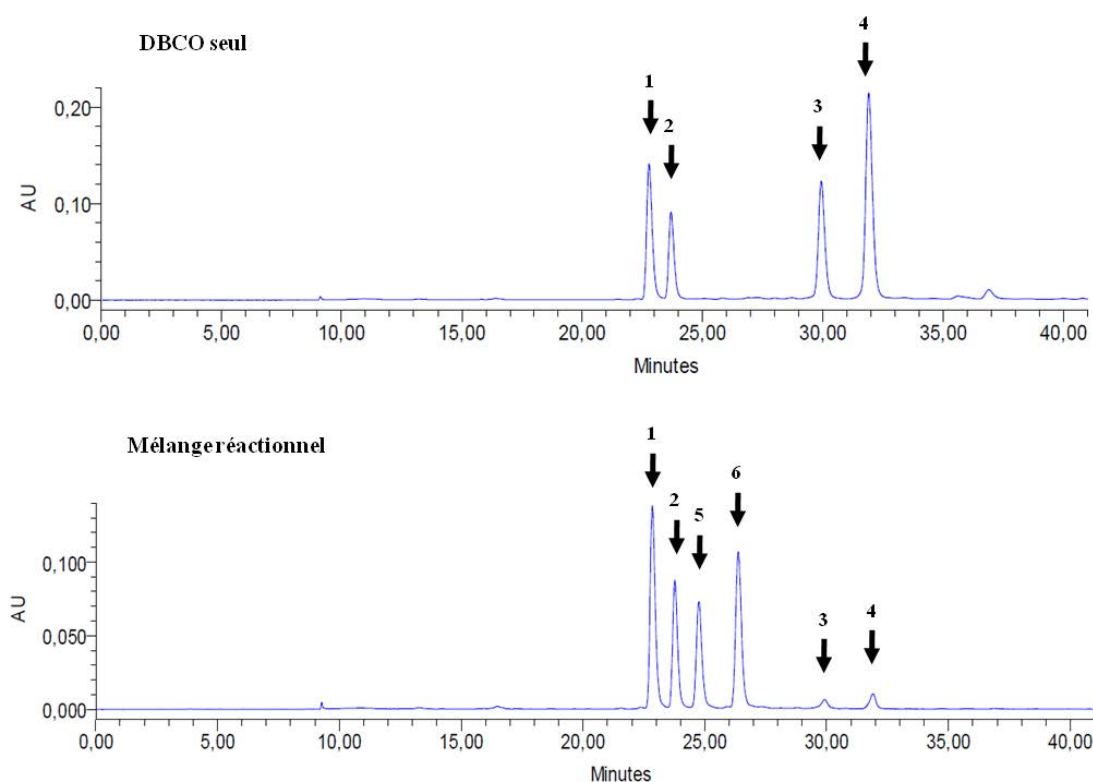


Figure 70. Chromatogrammes HPLC du DBCO-Alexa-Fluor 488 et du milieu réactionnel après 2h à 37°C (Détection à 488 nm).

2.1.4 Obtention de la BGT

La β -glucosyltransférase du phage T4 a été étudiée et cristallisée en présence d'UDPG il y a quelques années par l'équipe du Dr. Solange Moréra (LEBS, Gif sur Yvette)¹⁷⁷. Dans le cadre de notre collaboration, elle nous a fourni des échantillons de l'enzyme et nous avons également été amenés à produire celle-ci dans des bactéries *E. coli* transformées permettant sa surexpression et à la purifier nous même dans son équipe selon un protocole qu'elle avait mis au point¹⁷⁷. Il s'agit d'une purification en deux étapes : tout d'abord, une purification en fonction de la charge, par chromatographie sur résine échangeuse de cations, puis une purification en fonction de la taille, par chromatographie d'exclusion stérique (Figure 71).

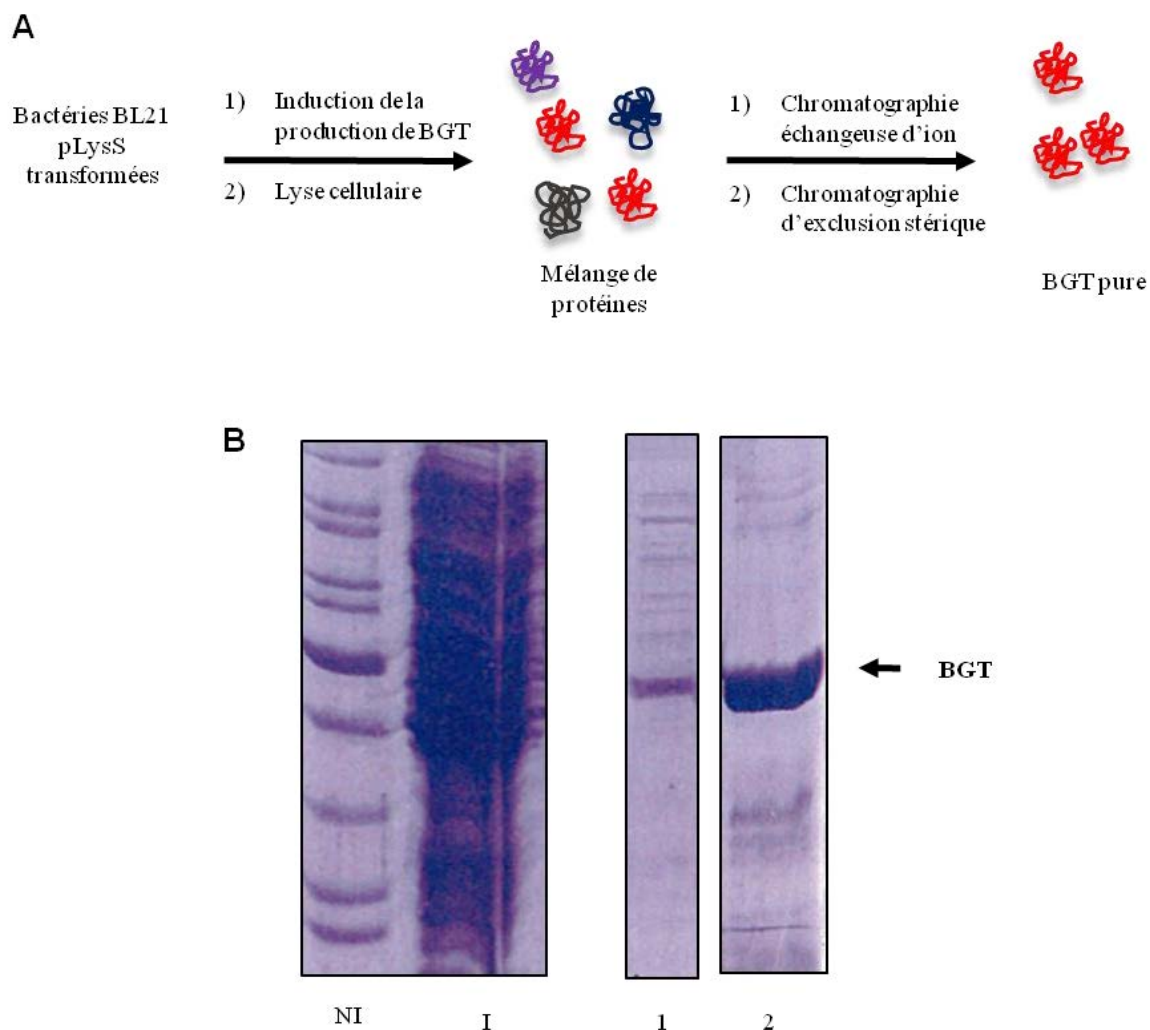


Figure 71. Production et purification de la BGT. A) Principe de la purification de la protéine BGT recombinante. B) Analyse de la purification par gel SDS-PAGE (NI : bactéries transformées où la production de BGT n'a pas été induite ; I : bactéries transformées où la production de BGT a été induite ; 1 : protéine après la première purification ; 2 : protéine après la seconde purification).

Les travaux de Chuan He utilisant l'UDP-6-azido-glucose¹⁴⁹ montraient que la BGT incorporait plus lentement le substrat modifié que son substrat naturel. Il nous a donc paru intéressant d'obtenir des informations sur l'interaction entre la BGT et l'UDP-6-azido-glucose qui pourrait nous permettre d'optimiser le transfert du glucose modifié par une BGT modifiée.

Disposant de données cristallographiques pour la BGT en présence de son cofacteur UDPG, nous avons donc regardé de près le site actif de l'enzyme et nous avons constaté que l'alcool présent en position 6 du sucre était à proximité (2.75 angströms) du résidu Glu22 de la poche catalytique (Figure 72). Le groupement azoture étant plus volumineux que le

groupement alcool, il est possible qu'une gêne stérique avec ce résidu soit à l'origine de la moindre efficacité du transfert du glucose modifié par rapport au glucose standard observée dans les travaux précédents. Solange Moréra nous a proposé de valider cette hypothèse en nous fournissant le mutant E22A qu'elle avait déjà préparé dans le cadre de travaux antérieurs, afin d'évaluer l'activité de ce mutant, en espérant que le remplacement par un acide aminé moins encombré améliorerait l'activité de l'enzyme.

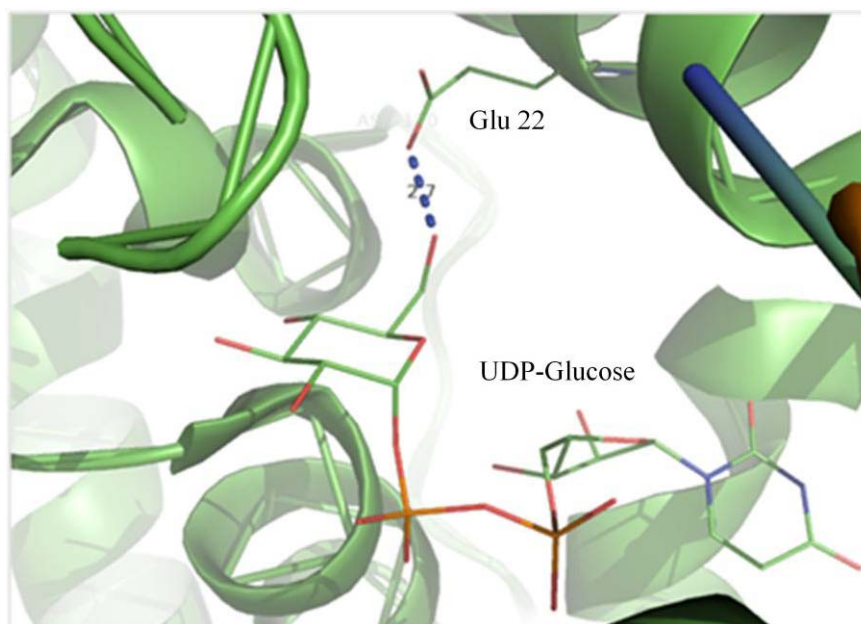


Figure 72. Représentation de la poche catalytique de la BGT faisant apparaître l'UDP-glucose et la distance entre l'alcool en position 6 du glucose et le résidu Glu 22.

En parallèle, nous avons fourni à l'équipe de Solange Moréra un échantillon de notre UDP-6-azido-glucose **9** afin d'essayer d'obtenir des cristaux de la BGT avec ce substrat modifié qui pourraient nous guider quant aux mutants appropriés à préparer. Malheureusement, malgré l'obtention de cristaux (Figure 73), aucune structure aux rayons X n'a pu être obtenue.

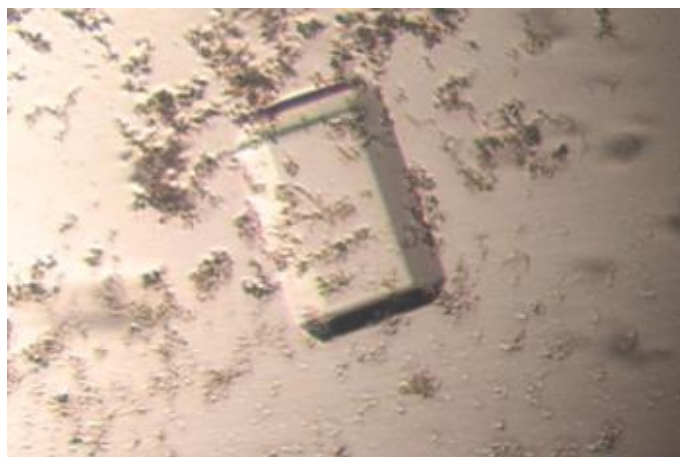


Figure 73. Cristal d'un mutant inactif de la BGT (D100A) en présence d'UDP-6-N₃-glucose 9.

2.1.5 Tentatives de transfert du glucose modifié sur le substrat modèle

Une fois obtenus les outils nécessaires à la mise en place de la méthode de fonctionnalisation d'ADN comportant les bases 5hmC, nous avons cherché à évaluer l'efficacité du transfert du glucose modifié sur notre substrat modèle, le duplex d'ADN présenté dans le paragraphe **2.1.1**. Nos différentes tentatives sont résumées dans le tableau 2.

Essai	[Duplex]	Température	Temps	[UDP-6-N ₃ -Glucose]	[BGT]	Tampon
1	30 μ M	37°C	2h	150 μ M	0.15 μ M	Tris-HCl 100 mM pH=7,9, 25 mM MgCl ₂
2	30 μ M	37°C	2h	150 μ M	0.15 μ M	Tris-HCl 100 mM pH=7,9, 25 mM CaCl ₂
3	30 μ M	37°C	2h	150 μ M	0.15 μ M	HEPES 50 mM pH=7,9, 25 mM MgCl ₂
4	30 μ M	37°C	4h	150 μ M	0.15 μ M	HEPES 50 mM pH=7,9, 25 mM MgCl ₂
5	30 μ M	37°C	4h	300 μ M	1.5 μ M	HEPES 50 mM pH=7,9, 25 mM MgCl ₂
6	30 μ M	37°C	48h	1 mM	30 μ M	HEPES 50 mM pH=7,9, 25 mM MgCl ₂

Tableau 2. Récapitulatif des différentes tentatives de transfert du sucre modifié sur le duplex modèle.

Quelles que soient les conditions, nous n'avons pas pu observer le transfert du glucose modifié par l'enzyme sauvage ou par le mutant E22A, malgré l'emploi de différentes techniques d'analyses (HPLC, LC-MS/MS, MALDI-TOF). Nous avons aussi tenté d'effectuer la réaction de transfert du glucose modifié dans les conditions 2 suivie de la fonctionnalisation par l'Alexa-Fluor 488. Après isolement du duplex par une colonne d'exclusion stérique Biospin P6, nous n'avons pas pu détecter de signal fluorescent.

Comme nous l'avons indiqué plus haut, nous nous sommes rendu compte au cours de cette étude que notre oligonucléotide comportant la 5hmC était en fait partiellement déprotégé ce qui nous a conduit à revenir sur ces essais ultérieurement, une fois ce problème de déprotection résolu, et à renouveler nos tentatives de glycosylation mais sans succès. La réaction en présence du substrat naturel, l'UDP-Glucose, conduite en parallèle, ne nous a pas permis de quantifier clairement le transfert du glucose mais nous avons toutefois observé en MALDI-TOF la présence d'un élargissement du signal de l'oligonucléotide avec une augmentation de masse de +160 en accord avec le transfert du glucose.

2.1.6 Conclusion

Nous avons donc tout le long de cette approche été confrontés à différents problèmes au niveau de la synthèse, de la détection mais aussi au niveau de la caractérisation de ces oligonucléotides fonctionnalisés. Le transfert de l'UDP-6-azido-glucose sur le duplex comportant la 5hmC n'a jamais pu être observé alors que l'approche a été depuis, de nombreuses fois utilisée dans la littérature. Ainsi, très récemment, un groupe a publié la méthode que nous souhaitons développer¹⁸². La réaction de l'azoture avec un groupement fluorescent relié à une dibenzylcyclooctyne permet le marquage spécifique des 5hmC dans des échantillons d'ADN génomique (Figure 74). Même si la réaction de chimie « click » n'a qu'un rendement de 66%, il est suffisant pour assurer une détection rapide et précise des 5hmC dans les échantillons d'ADN, jusqu'à 0.0002% du total de nucléosides.

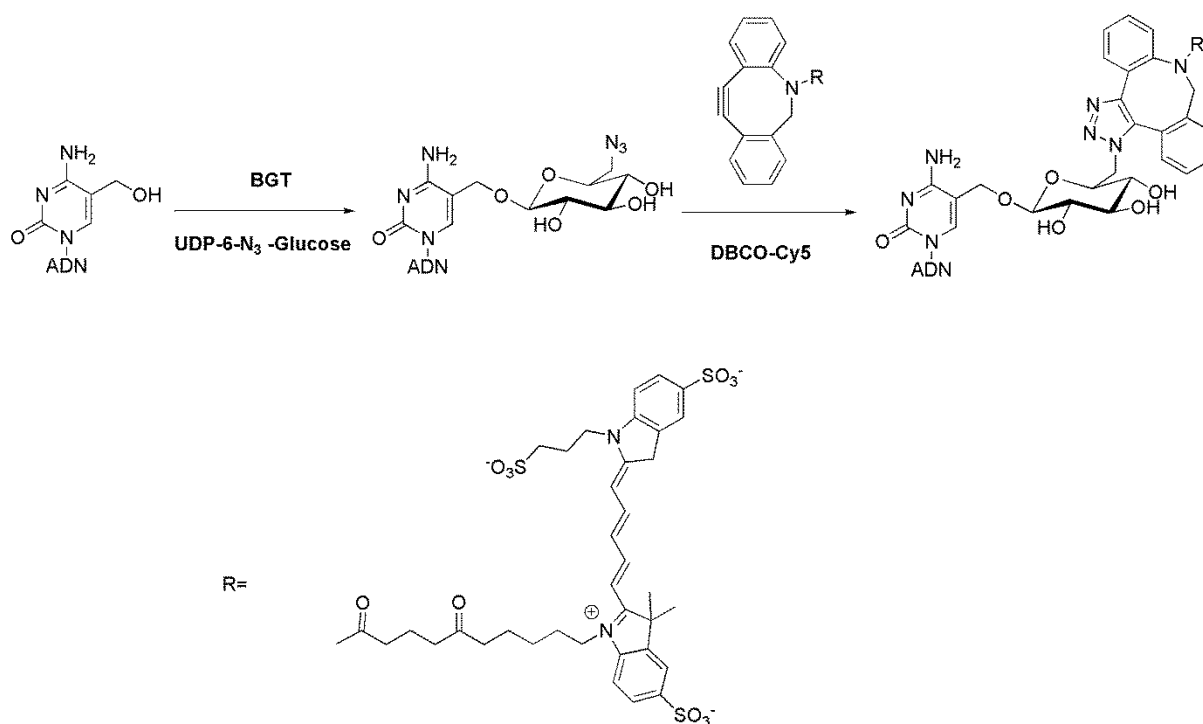


Figure 74. Fonctionnalisation des 5hmC par pour la quantification par fluorescence.

Notons également que d'autres équipes avec qui nous avons eu l'occasion d'échanger lors d'un congrès, ont été confrontées aux mêmes problèmes que nous dans des approches similaires de transfert de glucose modifié (S. Klimasauskas, résultats non publiés). Il faut également souligner qu'actuellement, aucune donnée ne précise clairement si la BGT effectue

la glycosylation sur n'importe quelle séquence et il est possible qu'il y ait eu dans notre cas un effet de séquence, ce qui pourra être confirmé en utilisant une séquence publiée dans la littérature.

2.2 Marquage chimique des 5hmC

Comme je l'ai mentionné plus haut, lorsque j'ai commencé ce travail, aucune étude rapportant la fonctionnalisation chimique des 5hmC dans un contexte ADN n'avait été proposé. Il nous a paru intéressant de nous y intéresser car une telle méthode pouvait présenter l'avantage par rapport à la méthode enzymatique de ne pas être spécifique de la séquence.

Au moment où peu de chose était encore connu sur la 5-formylcytidine dans l'ADN, nous nous sommes proposés d'essayer d'oxyder sélectivement la fonction hydroxyméthyle conduisant ainsi au motif 5-formylcytidine qui pourrait alors être marqué par une sonde comportant une fonction hydrazine (Figure 75).

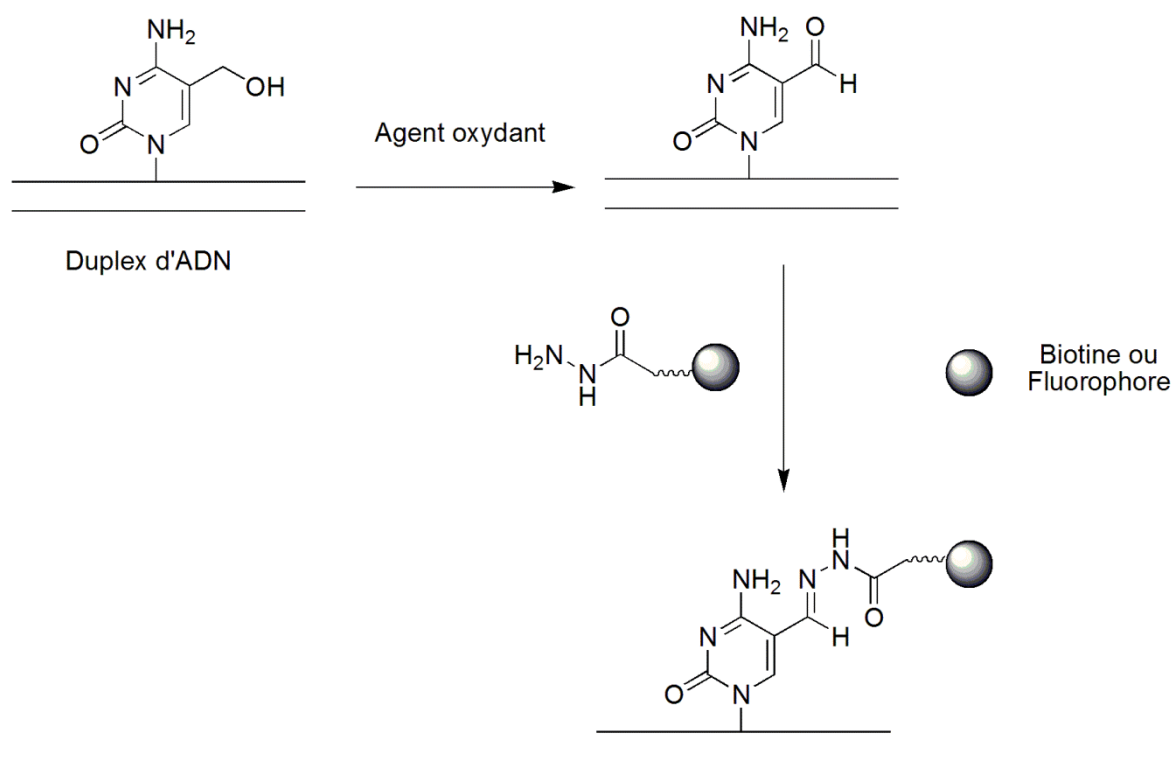


Figure 75. Oxydation sélective de l'alcool allylique de la 5hmC et fonctionnalisation de la 5fC formée.

Cette réaction, conduisant à la formation d'une hydrazone est couramment utilisée pour marquer les sucres dans les glycoprotéines, et de telles sondes biotinylées ou fluorescentes sont disponibles commercialement. Par ailleurs, des méthodes de dérivation de la 5-formyl-2'-désoxyuridine, une modification oxydative connue des thymidines, ont été rapportées dans la littérature¹⁸³. Ces méthodes, destinées à quantifier ces modifications, reposent sur la formation d'un conjugué hydrazone par réaction entre la fonction aldéhyde de la base oxydée et un réactif comportant une fonction hydrazine associée soit à un groupement fluorescent, soit à un ammonium quaternaire pour une détection plus aisée par spectrométrie de masse.

2.2.1 Mise au point des conditions d'oxydation et de fonctionnalisation de la 5hmC

Avant de mettre en œuvre cette approche sur des oligonucléotides modèles et au vu des problèmes de caractérisation rencontrés lors de notre approche enzymatique présentée dans le paragraphe précédent, il nous a paru d'abord nécessaire de faire la mise au point des conditions d'oxydation et de fonctionnalisation sur un modèle simple, le nucléoside 5-hydroxyméthyl-2'-désoxycytidine (5hmdC), ce que nous ne pouvions pas faire dans l'approche enzymatique, l'enzyme ayant besoin d'un substrat oligonucléotidique.

2.2.1.1 Synthèse de la 5-hydroxyméthyl-2'-désoxycytidine

Nous avons donc dans un premier temps réalisé la synthèse de la 5hmdC. Celle-ci a déjà été décrite et peut être envisagée à partir de différents nucléosides précurseurs, présentés dans la figure 76.

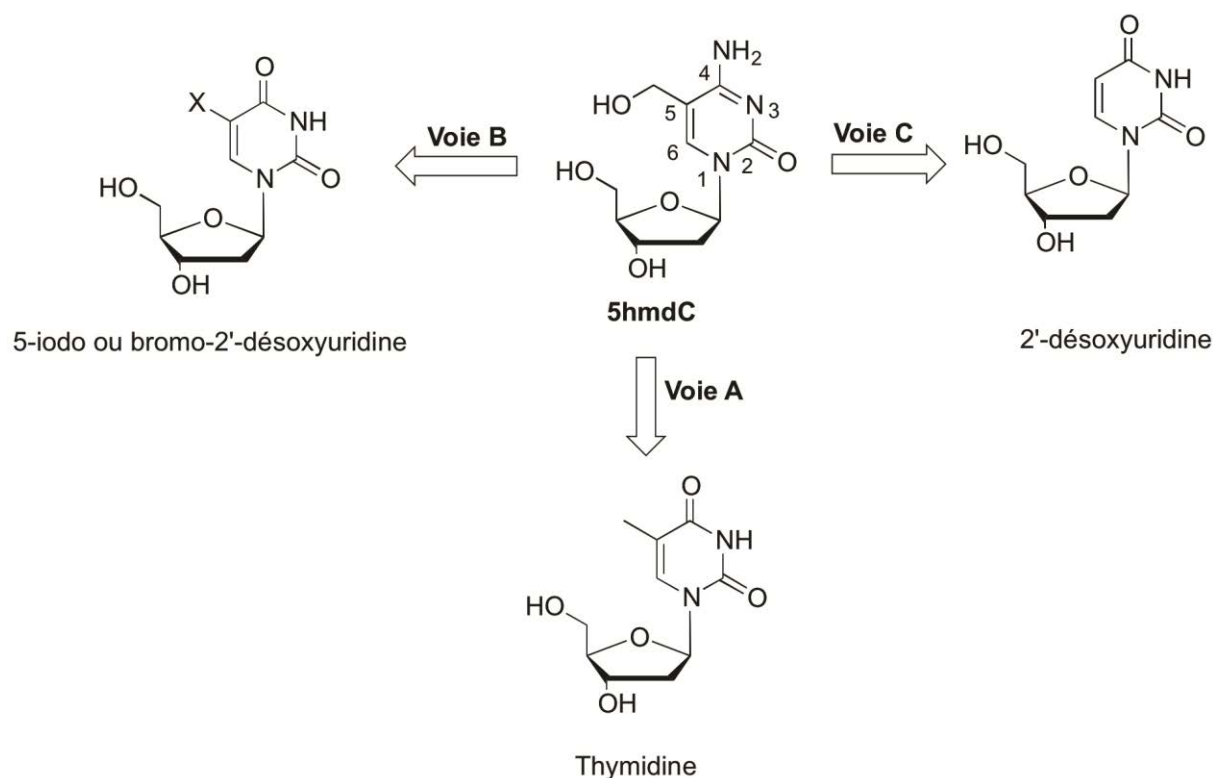


Figure 76. Voies rétrosynthétiques possibles pour obtenir la 5hmdC.

Dans tous les cas, le nucléoside de départ est sous la forme céto en position 4 du cycle pyrimidique et la synthèse nécessite une étape de conversion en fonction amine. Différents groupements protecteurs orthogonaux peuvent être employés pour la protection des fonctions hydroxyles selon que la molécule est destinée par la suite à être incorporée dans un oligonucléotide ou non. L'étape clé est donc l'introduction de la fonctionnalisation en position 5 de la base. Les nucléosides de départ peuvent donc être :

- Voie A : La thymidine, un nucléoside naturel qui peut être halogéné sur son méthyle en position 5 en présence d'Azobisisobutyronitrile (AIBN) et de *N*-Bromosuccinimide (NBS). L'halogène est ensuite substitué par le 3-hydroxypropionitrile, qui après déprotection conduit à l'alcool correspondant¹⁸⁴.
- Voie B : La 5-iodo ou 5-bromo-2'-désoxyuridine est commerciale et peut subir une formylation par l'action du monoxyde de carbone catalysée par du palladium. Le groupement formyle est ensuite réduit en alcool par $\text{NaBH}_4/\text{CeCl}_3$ ¹⁷⁹.

- Voie C : La 2'-désoxyuridine est également commerciale. L'hydroxyméthyle peut être introduit directement en présence de formaldéhyde¹⁸⁵. Nous avons choisi cette dernière voie qui nécessite moins d'étapes et des réactifs peu coûteux.

La 5hmdC a été donc synthétisée en 4 étapes, selon une procédure de la littérature (Figure 77)¹⁸⁵. La première étape consiste en l'hydroxyméthylation de la 2'-désoxyuridine (dU) en présence de formaldéhyde et de triéthylamine dans de l'eau, pour former le composé **11** avec un rendement de 60%. Les groupements alcools de celui-ci sont ensuite protégés par le TBDMSCl en présence d'imidazole dans le DMF pour donner le composé **12** avec un rendement non optimisé de 36%. La conversion du groupement carbonyle en position 4 du cycle pyrimidique est ensuite effectuée par un traitement au TPSCl dans l'acétonitrile en présence de DMAP et de triéthylamine, suivi d'un traitement à l'ammoniaque pour achever la formation du composé **13** avec un rendement de 65%. La déprotection finale des groupements TBDMS par le TBAF dans le dichlorométhane permet d'obtenir la 5hmdC avec un rendement 70%.

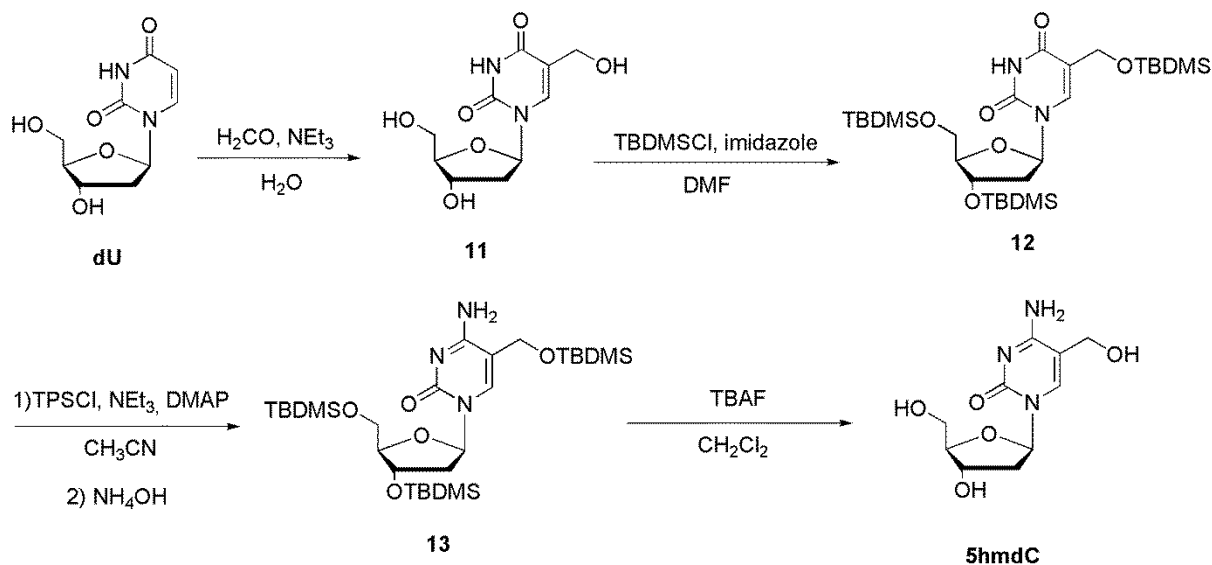


Figure 77. Synthèse de la 5-hydroxyméthyl-2'-désoxycytidine à partir de la 2'-désoxyuridine.

2.2.1.2 Mise au point des conditions d'oxydation de l'alcool allylique de la 5hmdC

Une fois la synthèse de la 5hmdC réalisée, nous avons effectué différents essais d'oxydation de l'alcool allylique afin d'accéder à la 5-formyl-2'-désoxycytosine (5fdC) (Figure 78), le but étant de définir des conditions de réactions simples compatibles avec un traitement d'échantillons d'ADN génomique. Pour cela, nous avons choisi des oxydants spécifiques des alcool allyliques et fait varier le solvant, la température et le temps de réaction, en nous appuyant sur des conditions décrites dans la littérature^{61,186,187}. Les différentes conditions utilisées ainsi que les rendements d'oxydation obtenus sont résumés dans le tableau 3.

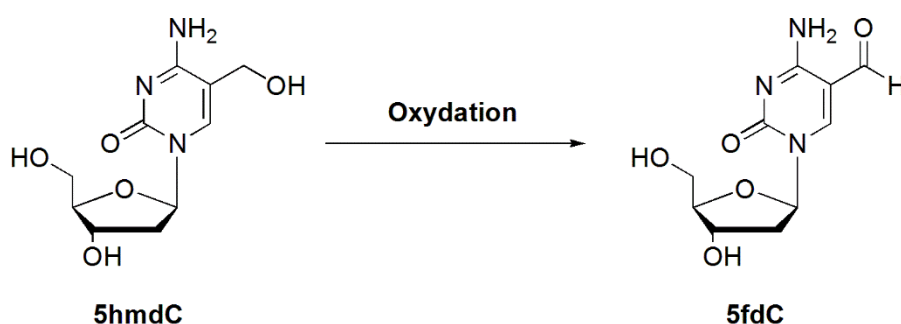


Figure 78. Oxydation sélective de la 5hmdC en 5fdC

Essai	Oxydant		Température	Solvant	Temps	Rendement
	Nature	Nombre d'équivalents				
1	RuO ₂	5	80°C	Eau/dioxane	9h	25 %
2	RuO ₂	5	RT	Eau/dioxane	12h	76 %
3	RuO ₂	5	60°C	Eau/dioxane	4h	90 %
4	RuO ₂	5	60°C	Eau	12h	80 %
5	Mn(OAc) ₃	6	RT	Eau/dioxane	24h	10 %
6	KRuO ₄	3	0°C	Eau, NaOH (0,05 M)	12h	50 %

7	KRuO₄	3	60°C	Eau, NaOH (0,05 M)	4h	67 %
8	MnO₂	6	RT	Eau/dioxane	12h	25 %
9	MnO₂	6	60°C	Eau/dioxane	4h	63 %

Tableau 3. Optimisation des conditions d'oxydation de la 5hmdC

Les réactions ont été suivies par CCM, puis, chaque brut réactionnel a été analysé par HPLC sur une colonne C18 en phase inverse (Figure 79). Les produits majoritaires ont été récoltés et analysés par spectrométrie de masse à la plateforme de Protéomique et de Spectrométrie de Masse de l'UPMC avec l'aide de Thierry Blasco.

L'analyse par HPLC et par spectrométrie de masse a permis d'établir que la quasi-totalité de la 5hmdC est convertie en 5fdC par l'action du dioxyde de ruthénium en 4 heures, dans un mélange eau/dioxane chauffé à 60°C (conditions 3). Avec l'oxyde de magnésium, le perruthénate de potassium ou l'acétate de manganèse, le produit de départ 5hmdC reste présent en grande quantité et il y a formation de produits secondaires que nous n'avons pas réussi à identifier en masse. Le rendement est quelque peu modifié par l'augmentation du temps de réaction, les différences ne sont pas significatives. Alors que le perruthénate de potassium a été très utilisé dans un certain nombre d'études pour oxyder la 5hmC, comme nous l'avons présenté dans la partie introductive de ce chapitre, nous avons pu observer que ce n'était pas l'oxydant le plus efficace dans les conditions que nous avons testées. La température idéale se situe pour l'ensemble des oxydants autour de 60 °C.

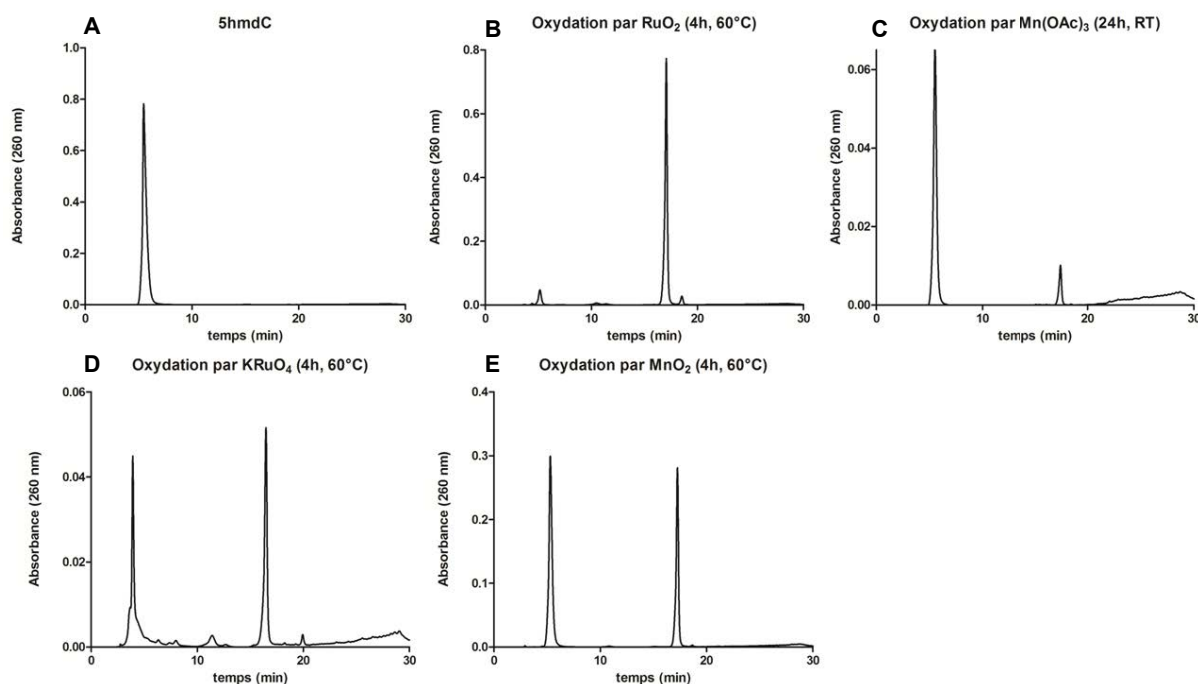


Figure 79. Chromatogrammes HPLC de **A.** la 5hmdC ; des bruts réactionnels de l'oxydation de la 5hmdC par **B.** RuO_2 (conditions 3), **C.** $\text{Mn}(\text{OAc})_3$ (conditions 5), **D.** K_2RuO_4 (conditions 7), **E.** MnO_2 (conditions 9).

Afin de s'assurer que l'oxydation de la 5hmC ne dégradait pas les autres bases de l'ADN, la 2'-désoxyadénosine (dA), la 2'-désoxyguanosine (dG), la 2'-désoxycytidine (dC) et la 2'-désoxythymidine (dT) ont été soumises aux conditions d'oxydation 3. L'analyse comparative par HPLC analytique des échantillons oxydés et non oxydés a permis de déterminer que l'oxydation par le dioxyde de ruthénium dans la condition 3 n'affectait que l'alcool allylique de la 5hmC et ne dégradait pas les autres nucléosides constitutifs de l'ADN.

2.2.1.3 Mise au point des conditions de fonctionnalisation de la 5hmdC par la biotine-hydrazine

L'étape de fonctionnalisation de la 5fdC obtenue précédemment a alors été réalisée en présence d'une biotine comportant une fonction hydrazine, en nous basant sur des conditions décrites dans la littérature, dans une solution aqueuse d'acide acétique à 10% à 37°C (Figure 80)¹⁸³.

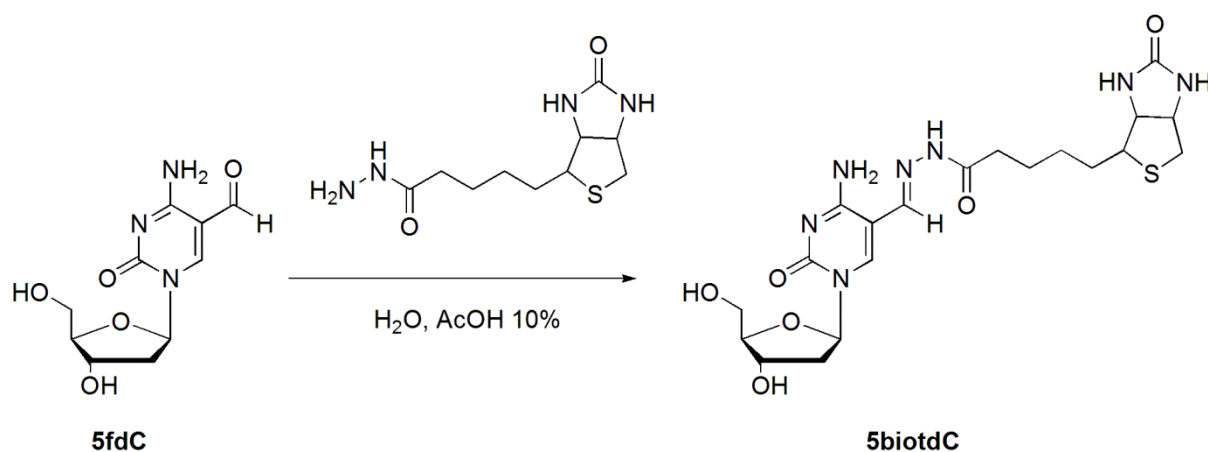


Figure 80. Fonctionnalisation de la 5fdC par la biotine hydrazine.

Le brut réactionnel a été analysé et purifié par HPLC sur colonne C18 en phase inverse (Figure 81). Les pics des produits ont été récoltés et analysés par spectrométrie de masse LC-MS/MS à la plateforme de l'UPMC comme précédemment. Au bout de 12 heures à 37°C, on observe la disparition totale de la 5fdC et la formation de trois produits dont deux majoritaires. La fraction 3 (majoritaire) correspond à la 5biotC (Figure 80 C). Les deux autres n'ont pas pu être identifiés.

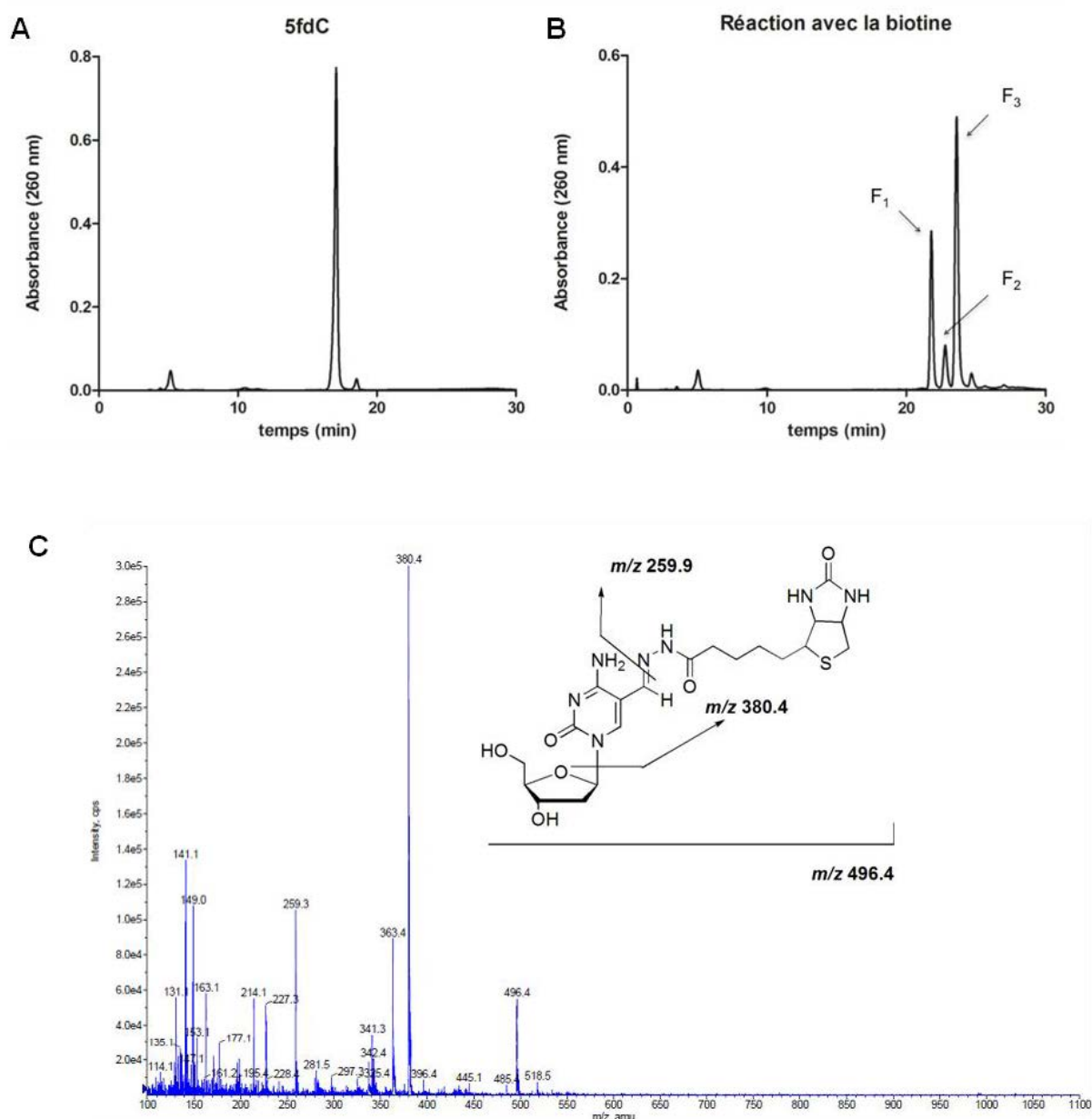


Figure 81. Fonctionnalisation de la 5fdC A) Profils HPLC de la 5fdC avant réaction ; B) Profils HPLC du brut réactionnel après 12 heures de réaction avec la biotine-hydrazine ; C) Spectre MS/MS de la fraction F3.

A l'issue de cette étude, nous avons mis au point sur des nucléosides modèles des conditions optimales d'oxydation de l'alcool allylique de la 5hmC pour obtenir la 5fC. La fonctionnalisation de celle-ci s'effectue ensuite simplement en présence d'hydrazine-biotine en milieu faiblement acide. Nous avons ensuite cherché à appliquer ces conditions à l'oxydation et à la fonctionnalisation d'oligonucléotides contenant la 5hmC.

2.2.2 Oxydation et fonctionnalisation d'un oligonucléotide modèle

Nous avons choisi comme séquence modèle le 30-mer présenté précédemment (5'-GAA GCT GGA CAG TA**5hmC** GTC AAG AGA GTG CAA-3'), comportant une 5hmC au milieu de la séquence, et nous l'avons soumis aux conditions présentées précédemment (Figure 82). Après l'oxydation, les échantillons ont été centrifugés pour éliminer le RuO₂ avant l'étape de fonctionnalisation.

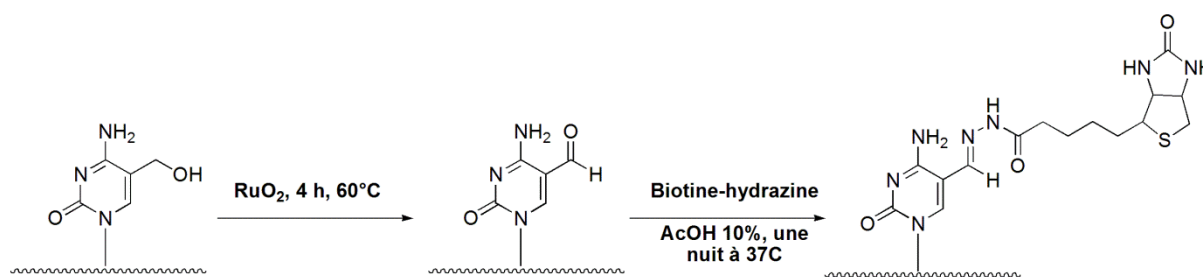


Figure 82. Oxydation et fonctionnalisation de l'oligonucléotide modèle.

Dans un premier temps, nous avons analysé cet échantillon ainsi qu'un contrôle non oxydé par LC-MS/MS après digestion enzymatique conduisant aux nucléosides constitutifs selon le protocole décrit précédemment dans le chapitre II. Les nucléosides constitutifs (dA, dC, dT, dG et 5hmdC) ont été détectés dans l'échantillon contrôle mais aucun signal n'a été détecté dans l'échantillon oxydé. Nous avons alors essayé de comprendre cet échec. Au cours de notre traitement, nous avons réalisé une étape de centrifugation après l'oxydation afin d'éliminer l'oxydant. Il semblerait que nous ayons perdu une grande partie de notre échantillon à cette étape ce qui expliquerait l'absence de signal au cours de cette première analyse. Nous avons confirmé cette hypothèse en mesurant l'absorption de notre échantillon à 260 nm sur un spectrophotomètre NanoDrop ; le surnageant après centrifugation n'absorbe quasiment plus et il est très probable que l'oligonucléotide soit resté complexé avec l'oxydant.

Nous avons essayé de résoudre ce problème en jouant sur plusieurs facteurs : (i) en variant les quantités et la nature de l'oxydant, (ii) en essayant de purifier l'oligonucléotide avant l'analyse, soit sur des petites colonnes d'exclusion stérique, soit, en faisant l'hypothèse que la fonctionnalisation par la biotine a fonctionné, par recapture et enrichissement des

oligonucléotides sur des billes de streptavidine. Les différentes tentatives sont résumées dans le tableau 4.

Essai	Concentration en oligonucléotide	Oxydant	Purification	Détection à 260 nm	Analyse en masse
1	5 μ M	RuO₂ 75 éq.	Aucune	Oui	LC/MS/MS : aucun nucléoside détecté
2	5 μ M	RuO₂ 75 éq.	Colonne d'exclusion	Non	Non déterminée
3	5 μ M	RuO₂ 75 éq.	Colonne d'exclusion	Non	Non déterminée
4	5 μ M	RuO₂ 75 éq.	Colonne d'exclusion	Non	Non déterminée
5	5 μ M	RuO₂ 75 éq.	Billes de streptavidine	Non	Non déterminée
6	5 μ M	RuO₂ 5 éq.	Billes de streptavidine	Oui	MALDI : aucun oligo détecté
7	5 μ M	KRuO₄ 100 éq.	Billes de streptavidine	Oui	MALDI : aucun oligo détecté
8	5 μ M	RuO₂ 5 éq.	Colonne d'exclusion	Oui	LC/MS/MS : aucun nucléoside détecté
9	10 μ M	KRuO₄ 10 éq.	Aucune	Oui	LC/MS/MS : nucléosides détectés
10	10 μ M	RuO₂ 10 éq.	Aucune	Oui	LC/MS/MS : nucléosides détectés

Tableau 4. Récapitulatif des différents essais de fonctionnalisation et de purification sur l'oligonucléotide modèle.

Nous avons d'abord cherché à purifier le mélange réactionnel à l'aide de petites colonnes d'exclusion stérique (essais 2 à 4). L'utilisation de ces colonnes permet d'éluer directement l'oligonucléotide, de masse élevée, tandis que les petites molécules ainsi que l'oxydant restent dans la colonne. Un contrôle a été réalisé avec un oligonucléotide standard, ne comprenant pas de base modifiée. La mesure des concentrations indique que si la quasi-totalité du contrôle est récupérée, l'oligonucléotide soumis aux conditions d'oxydation n'est pas présent en sortie de colonne. Le traitement oxydant semble donc perturber la purification par exclusion. Nos essais ayant été réalisés en large excès de RuO₂, nous avons essayé de diminuer les quantités par rapport aux faibles concentrations d'oligonucléotide sur lesquelles nous travaillons (essai 8) et nous avons pu éluer une partie de l'oligonucléotide mais l'analyse

par LC-MS/MS après digestion enzymatique n'a pas permis d'observer les nucléosides constitutifs.

Nous avons donc essayé une autre méthode pour récupérer l'oligonucléotide : après l'oxydation et la fonctionnalisation par la biotine-hydrazine, l'oligonucléotide doit comporter, si la réaction s'est bien déroulée, un groupement biotine qui peut permettre la purification de l'oligonucléotide par les billes magnétiques recouvertes de streptavidine. Le principe repose sur l'affinité très forte entre la biotine et la streptavidine ce qui permet de capturer uniquement les espèces biotinylées et par différentes étapes de lavage, d'éliminer les autres composés, puis après relargage en présence d'une solution à 0.1% en TFA, de récupérer les oligonucléotides biotinylés. Nous avons d'abord vérifié que notre protocole de capture sur billes de streptavidine et de relargage fonctionnait bien à l'aide d'un oligonucléotide commercial similaire déjà biotinylé à son extrémité 3' disponible au laboratoire. Nous avons ensuite tenté de purifier l'oligonucléotide ayant été soumis à un excès de RuO_2 puis aux conditions de fonctionnalisation (essai 5) : dans ce cas, nous n'avons pas observé de relargage d'oligonucléotide (essai 5). En présence de concentrations moindres en oxydant, un contrôle des concentrations au spectrophotomètre à 260 nm semblait indiquer que l'oligonucléotide était bien relargué (essai 6). Cependant, ni les analyses par MALDI-TOF ou par LC-MS/MS n'ont permis de mettre en évidence l'oligonucléotide biotinylé. Nous avons également à nouveau essayé d'utiliser le perruthénate de potassium même si nous avons vu dans nos essais préalables que l'oxydation était moins performante. Là-aussi, les essais de purification sur bille n'ont pas donné de résultats satisfaisants (essai 7).

Cependant, l'analyse directe (sans étape de purification) par LC-MS/MS des oligonucléotides oxydés par KRuO_4 , fonctionnalisés, puis digérés (essais 9 et 10) a cependant permis d'identifier la présence en plus des nucléosides constitutifs de 5fdC même si la 5hmdC n'avait pas entièrement disparu. L'oxydation en présence de KMnO_4 a donc bien eu lieu mais le rendement n'est pas quantitatif, ce que nous avons d'ailleurs aussi observé sur notre modèle nucléosidique. En revanche, nous n'avons pas détecté le nucléoside biotinylé : soit la réaction n'a pas eu lieu, soit elle s'est déroulée avec un rendement trop faible et nous sommes à la limite de détection, soit la digestion enzymatique n'est plus efficace quand le nucléoside est biotinylé. L'analyse MALDI de cet oligonucléotide n'a pas permis d'observer la fonctionnalisation par la biotine validant en partie la première hypothèse.

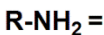
2.2.3 Conclusion

Nous avons donc montré la faisabilité de la fonctionnalisation des 5hmC par voie chimique à l'échelle nucléosidique et nous avons été confrontés à nouveau à des problèmes de détection et d'analyse lorsque nous sommes passés à l'échelle oligonucléotidique. Cependant, nous avons été en mesure d'observer l'oxydation partielle de la 5hmC dans l'oligonucléotide et, même si la fonctionnalisation par la biotine hydrazine n'a pu être observée sur l'oligonucléotide, il est très probable que nous soyons à la limite de détection pour observer cette fonctionnalisation.

Notons qu'au cours de ces travaux, une publication est parue proposant une méthode d'enrichissement des 5hmdC après digestion enzymatique de l'ADN et avant l'analyse par LC-MS/MS¹⁸⁸. Cette méthode, dont le principe est présenté sur la Figure 83, est basée sur la glucosylation de l'alcool allylique par la β -glucosyltransférase, suivie d'une purification des nucléosides fonctionnalisés par affinité. Nous pourrions donc envisager une application similaire à la suite réactionnelle que nous avons développée. Dans notre cas, l'ADN génomique serait tout d'abord digéré par le cocktail d'enzymes, oxydé puis fonctionnalisé et la 5hmC pourrait ainsi être enrichie par affinité avant l'analyse en masse.

Figure 83. Méthode d'enrichissement des 5hmdC avant analyse LC-MS/MS¹⁸⁸.

D'autres publications sont parues récemment montrant l'intérêt et la faisabilité de cette approche de marquage chimique pour la quantification de la 5hmC. La 5hmC est d'abord oxydée en 5fC par l'action du perruthénate de potassium, puis les fonctions aldéhydes formées vont réagir avec différentes molécules fluorescentes pour permettre la détection (Figure 84)^{189,190}. L'utilisation d'un polymère cationique qui va se fixer à l'ADN et servir de donneur d'énergie dans un transfert de fluorescence peut permettre d'améliorer grandement le signal fluorescent¹⁹¹.



138

Chapitre IV

Sondes photoactivables pour l'étude de protéines interagissant avec les bases rares de l'ADN

L'identification de protéines interagissant avec la 5hmdC dans l'ADN est cruciale afin de mieux comprendre les mécanismes biologiques associés. Les protéines qui se lient à l'ADN méthylé sont-elles les mêmes que celles qui se lient à l'ADN hydroxyméthylé ou existe-t-il des protéines de liaison spécifiques aux séquences hydroxyméthylées ? Il est possible de répondre en partie à ces questions en étudiant l'affinité relative de certaines protéines de liaison à l'ADN méthylé (les MBD) déjà connues et certains travaux mettent en lumière des différences de comportement, comme nous l'avons déjà évoqué dans l'introduction générale. Cependant, l'identification de protéines inconnues nécessite la mise en place de stratégies permettant de piéger les protéines affines pour les séquences comportant ces modifications et peu de travaux ont été réalisés à ce jour permettant d'atteindre cet objectif. C'est dans ce contexte que s'inscrivent les travaux présentés dans ce chapitre.

Afin de caractériser des protéines inconnues interagissant avec une sonde ADN, deux types d'approches peuvent être employées : des approches reposant sur les interactions non covalentes entre la sonde et les protéines d'intérêt ; c'est le principe des expériences de pull-down. Une autre stratégie est de provoquer la formation de liens covalents entre la sonde oligonucléotidique, contenant la base modifiée de l'ADN, et les macromolécules qui vont interagir spécifiquement avec elle. Pour parvenir à ce but, le photomarquage d'affinité constitue un outil de choix et c'est la stratégie que nous avons choisi d'adopter pour tenter de répondre à cette problématique.

Dans ce chapitre, nous présenterons donc les travaux que nous avons réalisés dans le cadre de la conception et l'évaluation de sondes photoactivables d'ADN destinées à la caractérisation de protéines affines pour la 5-hydroxyméthylcytosine. Dans une première partie, nous ferons un certain nombre de rappels bibliographiques concernant le principe du photomarquage d'affinité, les groupements photoactivables, et les méthodes pour les introduire dans des séquences oligonucléotidiques. Dans une seconde partie, nous présenterons la conception, la synthèse et l'évaluation de nos sondes. Enfin, nous discuterons des résultats obtenus et des perspectives concernant l'utilisation de nos sondes pour la caractérisation de protéines partenaires d'ADN hydroxyméthylé.

1 Rappels bibliographiques

Comme nous l'avons vu dans l'introduction générale, l'identification de protéines interagissant avec la 5hmC est d'une importance cruciale pour mettre à jour les différents processus biologiques dans lesquels cette base modifiée est impliquée. Afin de caractériser des protéines inconnues interagissant avec une sonde ADN, deux types d'approches peuvent être employées : des approches reposant sur les interactions non covalentes entre la sonde et les protéines d'intérêt ; c'est le principe des expériences de « pull-down » que nous avons évoquées dans le paragraphe 7.2.4 de l'introduction générale. Une autre stratégie est de provoquer la formation de liens covalents entre la sonde oligonucléotidique, contenant la base modifiée de l'ADN, et les macromolécules qui vont interagir spécifiquement avec elle. Pour parvenir à ce but, le photomarquage d'affinité, qui sera présenté dans le paragraphe ci-dessous, constitue un outil de choix.

1.1 Le photomarquage d'affinité

Le concept de photomarquage d'affinité (PhotoAffinity Labeling, PAL) a été introduit et développé dans les années 1960 par le groupe de Westheimer¹⁹². Depuis le développement de la protéomique et de techniques d'analyse par spectrométrie de masse de plus en plus sensibles et efficaces, le photomarquage d'affinité a suscité un regain d'intérêt considérable comme en témoigne la littérature abondante utilisant cette technique.

Cette méthode est basée sur l'utilisation des propriétés photochimiques de groupements dits photoactivables, pour étudier les interactions entre biomolécules. Dans cette méthode, un groupement photoactivable est introduit sur une molécule d'intérêt, nommée alors sonde. Suite à son exposition à un rayonnement ultraviolet, cette fonction stable dans des conditions normales va être activée et former transitoirement une espèce fortement réactive. Cette espèce réactive va pouvoir former une liaison covalente avec les macromolécules présentes à proximité de la molécule d'intérêt notamment les protéines (Figure 85).

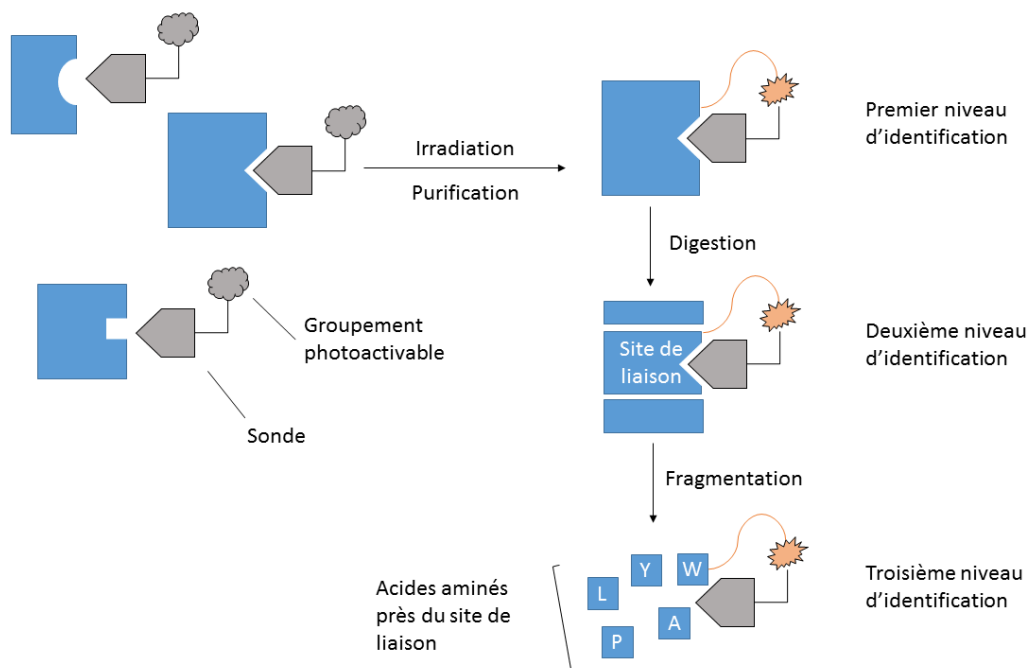


Figure 85. Principes du photomarquage et de l'identification des sites de liaisons. Adapté de 193

Cette technique permet d'accéder à plusieurs types d'informations : d'une part, elle peut permettre l'identification de protéines jusque-là inconnues interagissant avec la molécule d'intérêt ; d'autre part, elle permet d'obtenir des informations fines sur le site d'interaction protéine/molécule d'intérêt. Des étapes de purification et de caractérisation par spectrométrie de masse sont alors nécessaires pour accéder à ces informations.

D'autres fonctions supplémentaires peuvent être ajoutées sur les sondes photoactivables pour mieux exploiter la formation de ce lien covalent. L'ajout d'un groupement fluorescent ou d'un isotope radioactif peut par exemple faciliter la détection et l'identification des complexes covalents. L'ajout d'une biotine sur la sonde permet, grâce à l'interaction biotine/streptavidine, de purifier les complexes covalents formés dans des milieux biologiques complexes, notamment en vue d'une analyse ultérieure par spectrométrie de masse.

Le photomarquage d'affinité, couplé à différentes techniques d'analyses, a été particulièrement utilisé pour l'étude des interactions entre récepteurs-ligands, enzymes-inhibiteurs, enzymes-cofacteurs et protéines-acides nucléiques, ou encore protéines-membranes lipidiques¹⁹⁴. De nombreux travaux illustrent l'intérêt et la puissance de cette technique. Entre autres applications, on peut citer:

- L'identification de récepteurs : l'utilisation d'analogues de l'insuline radiomarquée et comportant un agent de photomarquage (4-azidobenzoyl) a permis d'identifier l'une des sous-unités du récepteur à l'insuline dans des extraits de membranes d'adipocytes¹⁹⁵.

- Caractérisation des cibles d'un médicament : l'utilisation de sondes photoactivables contenant des benzophénones a conduit à la découverte de nouveaux modes d'action de la vancomycine sur des bactéries¹⁹⁶. L'irradiation *in vivo* de *S. aureus* et de *E. faecalis* en présence de la vancomycine photoactivable a permis d'identifier deux nouvelles cibles de cet antibiotique, l'autolysine bifonctionnelle (ATL) et un transporteur ABC de peptide (pABC). Ces interactions ont ensuite été confirmées *in vitro* avec les protéines recombinantes.

- Etude de la structure, de la composition de la membrane et de ses partenaires protéiques : les premiers exemples de sondes lipidiques photoactivables ont été développés par le groupe de Khorana¹⁹⁷. Leurs travaux ont permis l'introduction de groupements diazo et azotures à différentes positions dans des lipides et la démonstration que ces sondes étaient capables de former des adduits covalents dans le contexte de vésicules. La synthèse de nombreuses autres sondes lipidiques photoactivables a depuis permis d'obtenir des informations sur les membranes biologiques. Par exemple, l'interaction spécifique entre la tubuline et un glycosphingolipide, le ganglioside GM1, a été mise en évidence par l'utilisation d'un analogue photoactivable et radioactif de ce ganglioside¹⁹⁸. La disparition de la radioactivité lorsque l'adduit tubuline-ganglioside était soumis à un traitement alcalin a permis d'identifier plus précisément le site de photomarquage, à savoir l'ancrage palmitoyl de la tubuline dans la membrane.

1.2 Les principaux groupements photoactivables

Pour mettre en œuvre une stratégie utilisant le photomarquage par affinité, il convient dans un premier temps de préparer la sonde comportant le groupement photoactivable le plus approprié. Ce dernier doit satisfaire à plusieurs critères :

- il doit être stable chimiquement dans les conditions expérimentales utilisées sur une gamme étendue de pH et stable et vis-à-vis de la lumière ambiante ;

- il ne doit pas changer drastiquement la structure de la molécule et entraîner le moins de gêne stérique possible afin de ne pas perturber l'affinité de la sonde pour sa cible ;
- il doit être activable à une longueur d'onde qui n'induirait pas de dégradation sur les autres molécules biologiques à étudier. Les protéines et les acides nucléiques absorbant en dessous de 300 nm, il doit être activé à une longueur d'onde supérieure à 300 nm ;
- suite à cette irradiation, il doit pouvoir générer des espèces fortement réactives, avec un temps de vie court pour éviter la dissociation du complexe d'intérêt et le photomarquage non spécifique ;
- il doit pouvoir former des liens covalents avec n'importe quel groupement fonctionnel présent dans les biomolécules, sans spécificité de fonction ou de liaison et notamment avec les liaisons C-H qui sont normalement inertes vis-à-vis de réactifs chimiques classiques ;
- les liens covalents formés doivent être suffisamment stables pour résister aux conditions expérimentales lors des analyses ultérieures ;
- ses propriétés photochimiques doivent être bien définies et donner lieu à un ou des adduits déterminés.

Depuis la première utilisation du para-nitrophenyl-diazoacetate pour photomarquer la chymotrypsine¹⁹², de nombreux groupements photoactivables ont été développés (Figure 86).

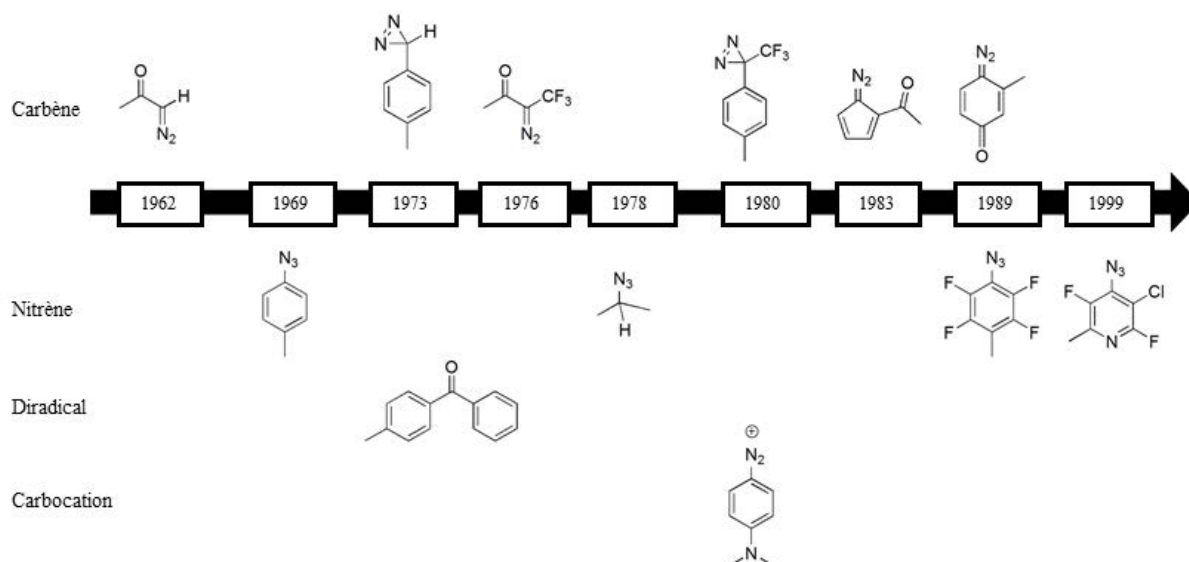


Figure 86. Date d'introduction des différents groupements de photomarquage. D'après¹⁹⁴.

Aucun d'entre eux ne s'est établi comme standard unique satisfaisant à tous les critères du photomarquage, mais trois fonctions principales se sont dégagées pour tout type d'application : les benzophénones, les diazirines et les azotures aromatiques (Figure 87).

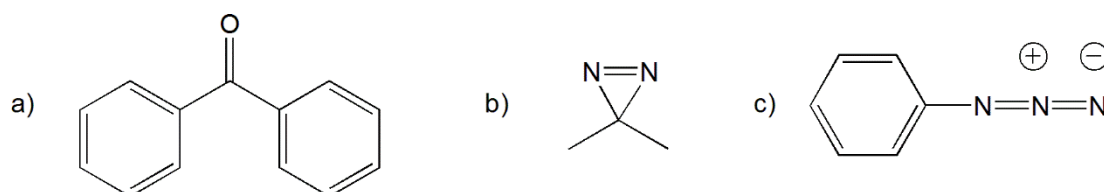


Figure 87. Groupements photoactivables les plus couramment utilisés ; (a) benzophénone ; (b) diazirine ; (c) azoture d'aryle.

Ces familles donnent lieu à la formation d'espèces réactives différentes : les benzophénones conduisent à la formation de radicaux ; les diazirines à des carbènes, et les azotures d'aryles à des nitrènes.

Pour ce qui concerne le cas particulier de l'étude des interactions protéine-acides nucléiques, des groupements photoactivables spécifiques de nature nucléosidique tels que la 5-Iodo-2'-désoxyuridine, la 5-Bromo-2'-désoxyuridine ou encore la 4-Thiothymidine ont été développés (Figure 88).

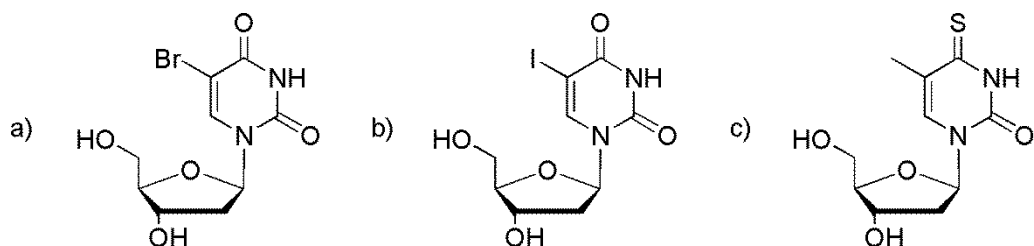


Figure 88. Groupement photoactivables nucléosidiques ; (a) 5-Bromo-2'-désoxyuridine (b) 5-Iodo-2'-désoxyuridine (c) 4-Thiothymidine.

La suite de cette introduction sera consacrée à la présentation des trois grandes familles de groupements photoactivables généraux ainsi que de ceux spécifiques aux acides nucléiques. Leurs avantages et inconvénients pour les applications de photomarquage d'affinité seront discutés, ainsi que leur photochimie et leurs voies de synthèse principales si les molécules ne sont pas accessibles commercialement.

1.2.1 Les azotures d'aryle

Les azotures d'aryle sont une des premières fonctions photoactivables utilisées, dès 1969, permettant l'identification de sites d'interactions d'un anticorps avec un haptène photoactivable¹⁹⁹. Ces groupements sont activés avec des longueurs d'ondes inférieures à 300 nm, entraînant la perte de diazote et la formation d'un nitrène singulet. Ce nitrène singulet **15** peut réagir *via* des mécanismes d'insertion dans des liaisons C-H ou X-H, où X est un hétéroatome (Figure 89). Cette espèce singulet a un faible temps de vie et est rapidement convertie en d'autres espèces réactives^{200,201}. Premièrement, le nitrène singulet va devenir un nitrène triplet **16** par un mécanisme de croisement intersystème (Intersystem Crossing, ISC). Cet état triplet **16** peut réagir avec des liaisons R-H par une abstraction d'hydrogène suivie de couplage avec le radical formé. La deuxième réaction possible est un réarrangement intramoléculaire rapide (entre 10 et 100 ps) qui aboutit à la formation d'une forme benzazirine **21** puis d'une forme dehydroazepine **22** par expansion de cycle²⁰². Ces deux composés, en équilibre, sont stables en solution et particulièrement réactifs vis-à-vis de nucléophiles. L'utilisation d'azotures d'aryles fluorés stabilise l'état nitrène singulet et permet de rendre la transformation en benzazirine puis dehydroazepine plus lente²⁰³. Ces composés sont donc plus avantageux puisque l'on souhaite favoriser la réactivité de la sonde photoactivable par insertion R-H.

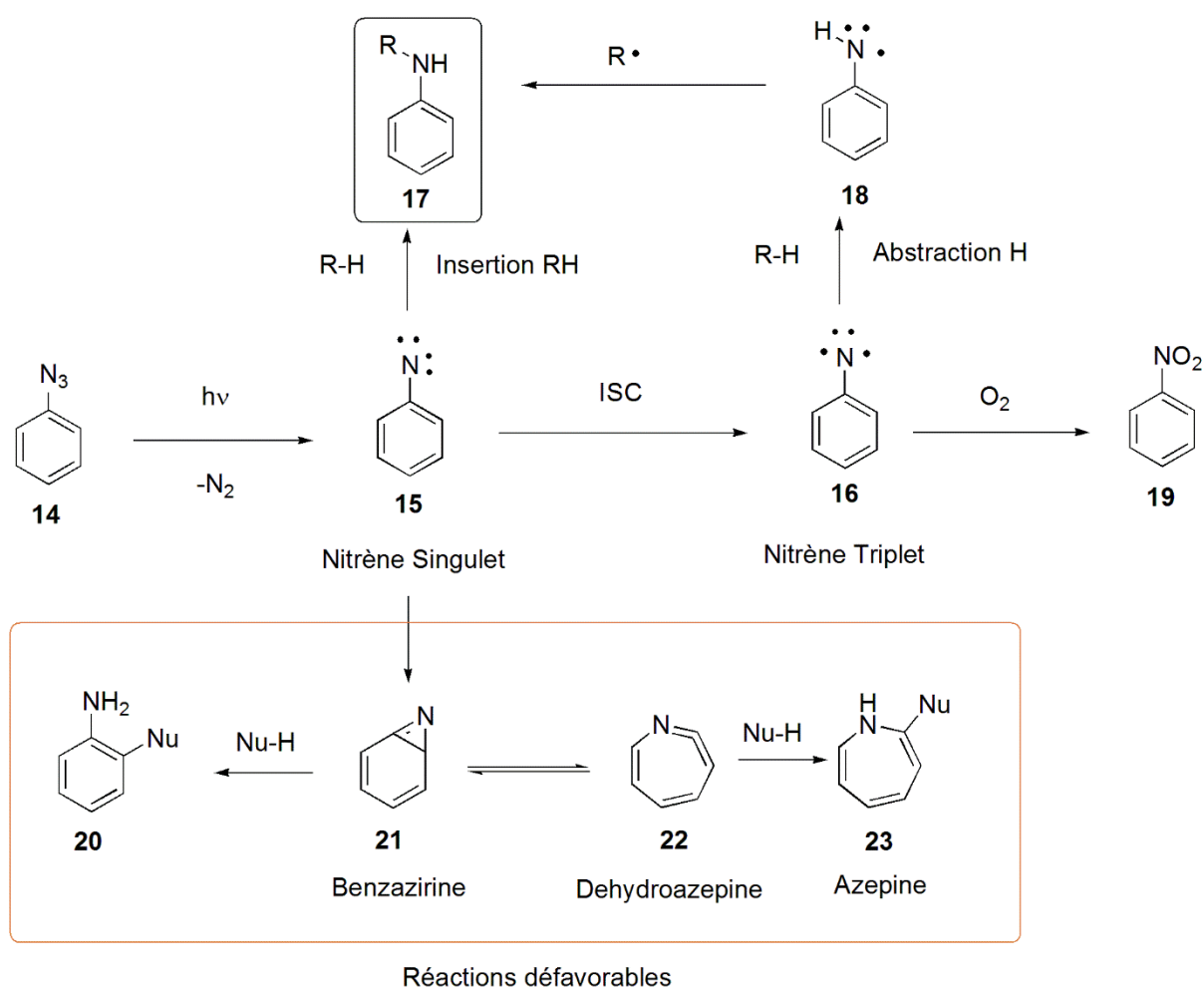


Figure 89. Principaux mécanismes de réaction des azotures d'aryles activés

Les principaux avantages des azotures d'aryles sont leur taille peu importante et leur structure qui permet de les incorporer facilement dans les molécules comportant des cycles aromatiques, comme par exemples la tyrosine, la phénylalanine ou bien les bases nucléiques (Figure 90). Leur facilité de synthèse en une ou deux étapes à partir des dérivés amino ou halogéno correspondants en a fait un des groupements photoactivables les plus utilisés lors des premières études utilisant le photomarquage.

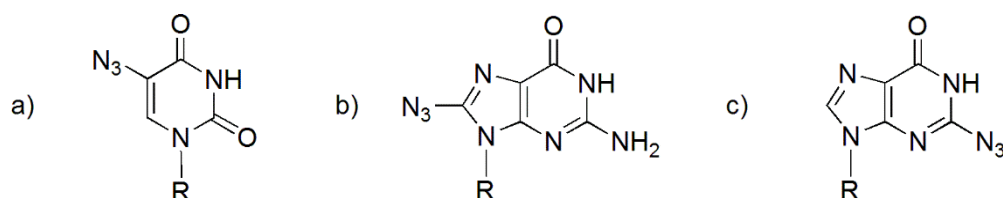


Figure 90. Différents analogues de bases de l'ADN contenant un groupement azoture ; a) 5-azidouridine ; b) 8-Azido-adenosine ; c) 2-Azido-adenosine.

Cependant, les longueurs d'ondes inférieures à 300 nm nécessaires pour activer les azotures d'aryles constituent un de leurs inconvénients majeurs, puisqu'elles peuvent endommager les protéines et les acides nucléiques présents en solution. La substitution du cycle aromatique par des groupements nitro, imino ou acyl permet de remédier partiellement à ce problème en augmentant la longueur d'onde maximale d'absorption²⁰⁴. Par ailleurs, les azotures d'aryles sont particulièrement sensibles aux réducteurs comme le dithiotréitol (DTT), qui les réduisent en donnant lieu aux amines aromatiques correspondantes. De plus, l'état triplet du nitrène **16** est aussi sensible à l'action du dioxygène qui peut le désactiver et former le composé nitro **19** correspondant. Les longueurs d'ondes d'irradiation nécessaires ainsi que les faibles rendements de photomarquage causés par les réactions secondaires font qu'on préfère si possible d'autres fonctions photoactivables aux azotures d'aryles.

1.2.2 Benzophénone

La benzophénone comme groupement photoactivable a été développée dans les années 1970, un des premiers exemples étant le photomarquage de l'albumine sérique bovine (BSA) par une hormone peptidique, la pentagastrine, fonctionnalisée avec des dérivés de benzophénone²⁰⁵. L'usage de la benzophénone s'est ensuite étendu dans les années 1980, notamment pour l'étude des interactions peptides-protéines, comme par exemple l'étude de l'interaction entre la substance P et son récepteur NK-1. Dans une série d'étude menée par le groupe de Leeman, des extraits membranaires de cellules CHO surexprimant les récepteurs de la substance P ont été ainsi irradiés en présence d'un analogue photoactivable et radioactif de la substance P. La combinaison de digestions enzymatiques par différentes protéases a permis d'identifier dans un premier temps la zone du récepteur interagissant avec le peptide²⁰⁶, puis dans un deuxième temps, grâce à l'analyse par spectrométrie de masse MALDI-TOF, l'acide aminé précis lié de façon covalente par le groupement benzophénone sur l'analogue de la substance P²⁰⁷.

Lors de son activation à une longueur d'onde comprise entre 350 et 365 nm la benzophénone **24** absorbe un photon, entraînant la formation rapide d'un di-radical **25** dans un état triplet via un croisement intersystème à partir de l'état excité singulet (Figure 91). Le di-radical à un temps de vie compris entre 80 et 120 μ s. S'il n'a pas réagi au bout de ce temps, l'état triplet **25** peut se relaxer pour revenir à l'état fondamental **24**. Le di-radical va réagir

avec une liaison X-H à proximité, tout d'abord en procédant à une abstraction d'hydrogène puis à la création d'une liaison C-X avec le radical carboné formé. L'abstraction de l'hydrogène est l'étape limitante, et est déterminée par la facilité à rompre de façon homolytique la liaison X-H et la stabilité du radical formé lors de l'abstraction. La benzophénone est ainsi plus réactive vis-à-vis des liaisons C-H que des liaisons O-H ou N-H²⁰⁸.

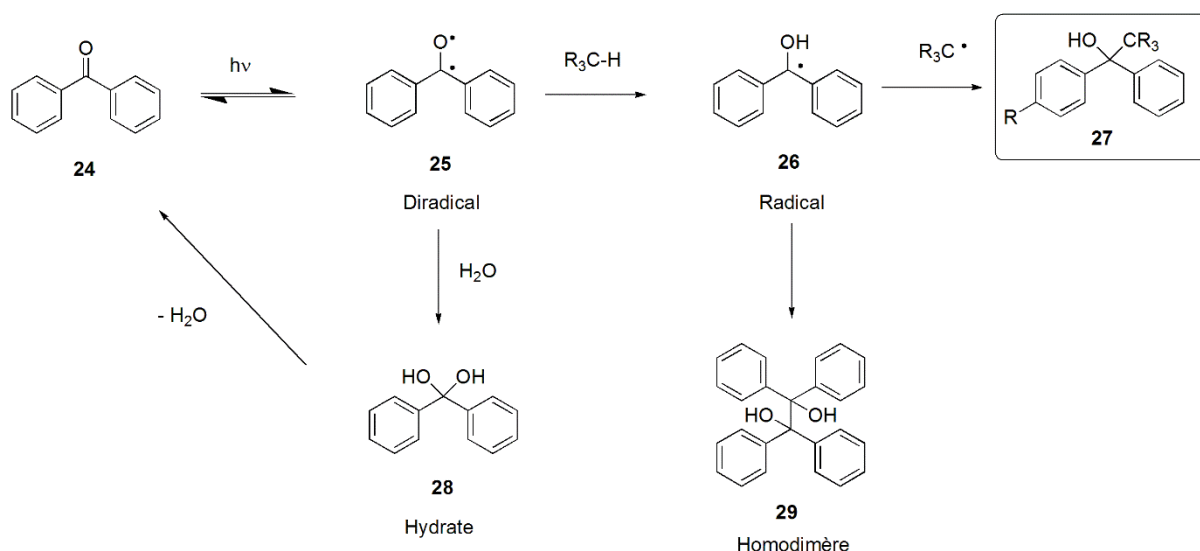


Figure 91. Principaux mécanismes de réaction de la benzophénone irradiée.

Le motif benzophénone comporte de nombreux avantages : il peut par exemple former des liaisons covalentes avec tous les acides aminés, même si la benzophénone a une plus grande affinité pour la méthionine²⁰⁹. Un autre avantage des benzophénones est que les réactions secondaires du di-radical **25** sont limitées. Il y a peu de réactions de réarrangements intramoléculaires. Si le di-radical réagit avec l'eau pour former l'hydrate **28** correspondant, cette espèce va rapidement se réarranger pour reformer la benzophénone **24** initiale. L'homodimérisation du monoradical est limitée par la proximité des deux radicaux formés lors de la rupture de la liaison C-H ainsi que la rapidité de la recombinaison qui suit. Un des principaux avantages de la benzophénone vis-à-vis des autres groupements photoactivables est liée à sa capacité à se régénérer. Au lieu d'être détruite définitivement lors de l'irradiation, la benzophénone n'ayant pas formé de liaison covalente peut ainsi revenir à l'état fondamental, non excité, et participer à d'autres cycles d'excitation/relaxation, permettant d'augmenter les effets du photomarquage au cours du temps.

Les deux désavantages principaux du groupement benzophénone sont sa taille et sa rigidité, qui peuvent entraîner une gêne stérique importante, ainsi que les temps d'irradiation nécessaires pour mener à bien le photomarquage, qui sont plus élevés que pour les autres groupements photoactivables. En ce qui concerne la synthèse de sondes comportant ce motif, de nombreux blocs commerciaux sont disponibles (comportant des substituants amine, acide carboxylique, halogène, maléimide, ...) permettant le couplage aisé sur tous types de composés.

1.2.3 Diazirine

Les diazirines sont des groupements photoactivables dont la longueur d'onde d'irradiation est comprise entre 330 et 380 nm. La première description des diazirines dans le cadre du photomarquage date de 1973, avec la synthèse et l'évaluation de la 3-aryle-3H-diazirine²¹⁰. L'optimisation de ces composés a rapidement conduit à l'utilisation des trifluorométhylaryldiazirines, qui sont les plus efficaces dans le cadre du photomarquage d'affinité et forment aujourd'hui la majorité des sondes diazirines utilisées.

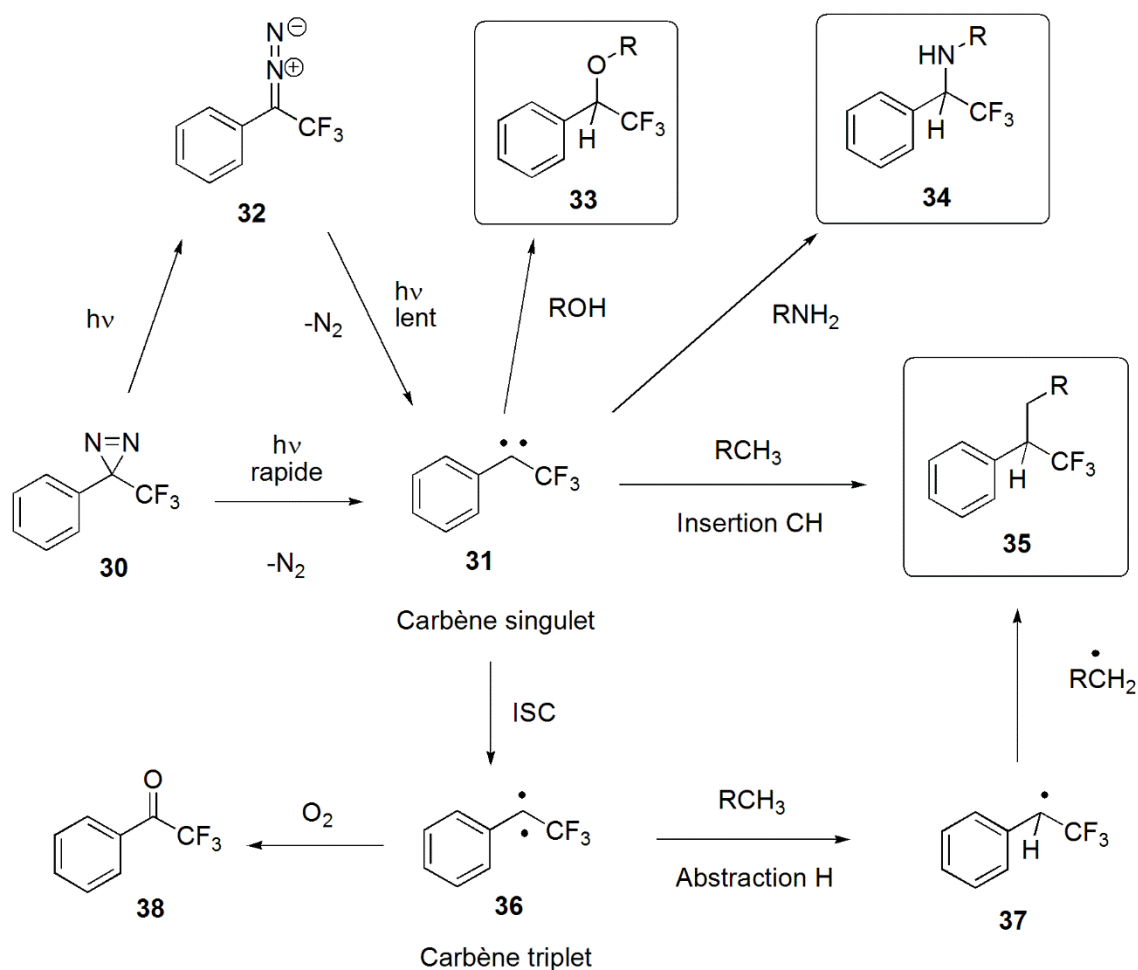


Figure 92. Mécanismes réactionnels de l'excitation de la trifluorométhylphényldiazirine.

L'irradiation de la molécule **30** va entraîner la perte d'une molécule de diazote et donner lieu à la formation d'un carbène singulet **31**, hautement réactif, avec un temps de vie très court (1 ns). Ce carbène singulet peut réagir en s'insérant dans de liaisons R-H (Figure 92). Le composé **31** réagit de façon préférentielle avec les liaisons O-H plutôt que les liaisons C-H, ce qui conduit à l'inactivation rapide des diazirines irradiées en milieux aqueux. Ce carbène singulet **31** peut aussi changer d'état de spin et devenir un carbène triplet **36** par conversion intersystème. Le carbène triplet va se comporter comme un di-radical et va pouvoir réaliser une abstraction d'hydrogène suivie de la formation d'une liaison covalente C-C.

Des réactions secondaires peuvent avoir lieu suite à l'irradiation. Premièrement, lors de l'irradiation de la diazirine **30**, il y a une photoisomérisation et formation du composé diazo **32** (Figure 92). L'irradiation de ce composé donne aussi lieu à la formation d'un carbène mais plus lentement que pour la diazirine²¹¹. Le composé **32** est aussi instable en

milieu acide et sensible aux additions nucléophiles, ce qui peut conduire à la formation d'adduits covalents non spécifiques. L'impact de cette réaction secondaire est limité quand on utilise des trifluorométhylaryldiazirines, le composé diazo **32** étant alors plus stable en milieu acide et moins sensible aux additions nucléophiles²¹².

Parmi les autres réactions secondaires possibles, on peut former, lors de la réaction du carbène singulet avec une amine secondaire ou primaire, une énamine **39** en équilibre tautomérique avec l'imine **40** correspondante (Figure 93). Ces tautomères peuvent ensuite être hydrolysés en milieux aqueux pour donner le composé **41** et rompre la liaison covalente préalablement formée entre la sonde photoactivable et la biomolécule²¹³. Le carbène triplet **36** peut aussi être détruit, soit en réagissant avec du dioxygène pour former une cétone, soit en réalisant deux abstractions radicalaires d'hydrogène successives.

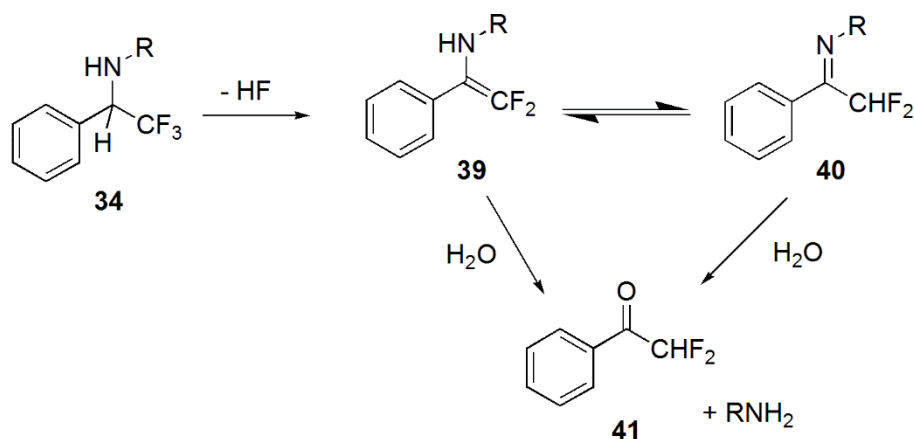


Figure 93. Réaction secondaire ayant lieu après l'insertion du carbène dans une liaison N-H.

Ces réactions secondaires, néfastes pour les rendements de photomarquage, font partie des désavantages des sondes diazirines. Même si l'utilisation de trifluorométhylaryldiazirines permet de limiter les effets de l'isomérisation en composé diazo, celui-ci est tout de même formé à hauteur de 35% de la diazirine d'origine. Le carbène singulet étant très réactif vis-à-vis des liaisons O-H, les diazirines vont de plus être très rapidement désactivées en milieux aqueux, ce qui peut diminuer le rendement de photomarquage.

Cette forte réactivité peut toutefois être aussi vue comme un avantage : l'inactivation rapide par les molécules d'eau va limiter drastiquement les réactions de photomarquage non spécifiques et permettre des temps d'irradiation rapides. Les autres avantages des diazirines comprennent leur stabilité vis-à-vis des additions nucléophiles et de la lumière, ainsi que des

conditions acides ou basiques. L'absorption à des longueurs d'onde largement supérieures à 300 nm est aussi, comme on l'a vu précédemment, un avantage non négligeable.

La synthèse des diazirines et notamment des trifluoromethylaryldiazirines a été bien développée et optimisée (pour une revue, voir ²¹⁴). La stratégie principale est présentée dans la Figure 94. A partir de la cétone correspondante et par action de l'hydroxylamine, on va former une oxime. La tosylation de cette oxime suivie d'un traitement à l'ammoniac conduit à la formation de diaziridines. L'oxydation finale permet d'obtenir les diazirines recherchées. Dans le cas d'applications nécessitant une sonde photoactivable avec peu d'encombrement stérique, on peut utiliser des diazirines aliphatiques.

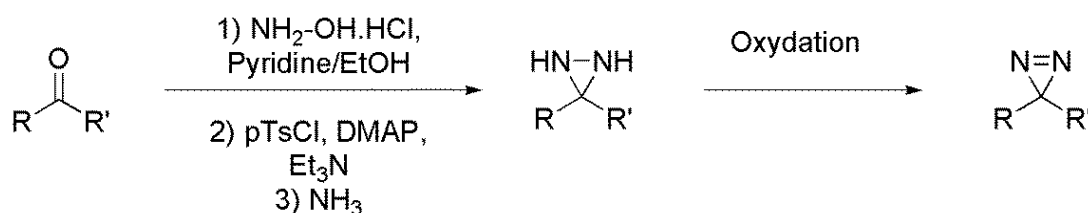


Figure 94. Voie principale de synthèse des diazirines.

1.3 Analogues de nucléosides

En dehors de ces groupements photoactivables utilisables pour de nombreuses applications, d'autres groupements ont été développés pour étudier les interactions mettant en jeu l'ADN. Ce sont des analogues de nucléosides naturels qui possèdent des propriétés photochimiques utilisables dans le contexte du photomarquage d'affinité.

1.3.1 Analogues thiolés d'acides nucléiques

La 4-Thiothymidine (^{4S}T), la 4-Thiouridine (^{4S}U) et la 6-Thio-2'-desoxyguanosine (^{6S}dG) sont des analogues des acides nucléiques dans lesquels un atome d'oxygène du cycle purique ou pyrimidique est remplacé par un atome de soufre (Figure 96). Cette substitution leur apporte une des propriétés importantes, mentionnée auparavant pour les sondes photoactivables : leur spectre d'absorption diffère de ceux des bases classiques, avec un

maximum autour de 330 nm ce qui permet de les activer avec des rayonnements qui ne vont pas endommager les acides aminés et les acides nucléiques (Figure 95).

Figure 95. Spectres d'absorption de la $^4\text{S}\text{T}$ et de la Thymidine (T). Tiré de ²¹⁵.

La $^4\text{S}\text{U}$ est un nucléoside rare, présente naturellement en un seul exemplaire dans les ARN de transfert (ARNt) d'*E. coli*^{216,216}. Cette occurrence naturelle a permis l'une des premières utilisations de la $^4\text{S}\text{U}$ dans du photomarquage d'affinité. L'irradiation d'un ARNt d'*E. coli* a permis d'obtenir des informations sur la structure de celui-ci en utilisant la $^4\text{S}\text{U}$ présente à la position 8 comme sonde structurale intrinsèque²¹⁷. Elle n'est pas présente dans les ARNt des cellules eucaryotes. La $^4\text{S}\text{T}$ et la $^6\text{S}\text{dG}$ ne sont pas des nucléosides naturels mais peuvent être obtenus par synthèse chimique. L'une des méthodes consiste à former le dérivé triazolo qui, après réaction en présence d'acide thiolacétique, conduit au nucléoside en version thiolée²¹⁸.

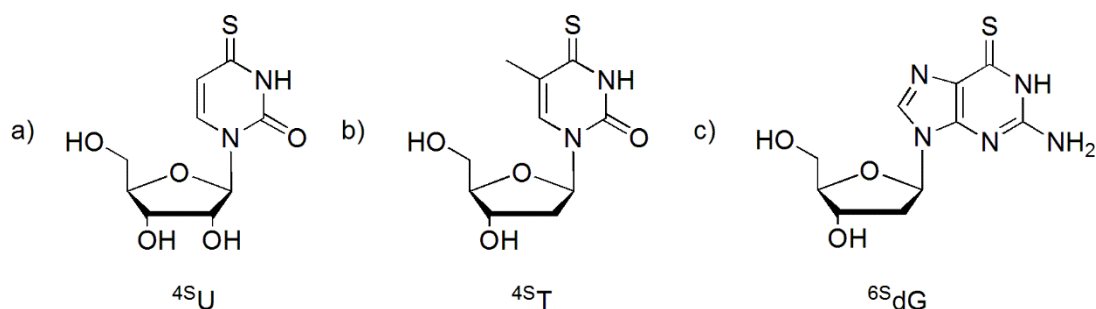


Figure 96. Analogues thiolés des bases de l'ADN. a) 4-Thiouridine ; b) 4-Thiothymidine c) 6-Thio-2'-désoxyguanosine.

La $^4\text{S}\text{T}$, $^4\text{S}\text{U}$ et la $^6\text{S}\text{dG}$ ont des comportements photochimiques similaires. L'absorption d'un photon permet de passer de l'état fondamental S_0 à un état excité singulet, S_2 . En quelques picosecondes, cet état singulet va se transformer de manière quasi-totale en état triplet T_1 par un mécanisme de conversion intersystème²¹⁹. C'est cet état triplet, avec un temps de vie compris entre 0.1 et 1 μs , qui est l'espèce réactive et va établir des liaisons covalentes avec les molécules présentes à proximité. La photochimie des dérivés thiolés est moins

maîtrisée que celle des groupements photoactivables présentés précédemment. Le mécanisme par lequel la ^{45}T établit des liaisons covalentes n'est pas entièrement élucidé, mais elle est capable de former des photoadduits avec tous les acides nucléiques standards²²⁰. La ^{45}T a été utilisée dans de nombreuses applications de photomarquage d'affinité sur des protéines, mais la nature des adduits et des acides aminés impliqués n'est pas connue. Le seul adduit caractérisé en solution entre les dérivés thiolés et un acide aminé est celui entre la 1,3-diméthyl-4-thiouridine et la L-lysine. L'irradiation conduit à une addition nucléophile de l'azote de la lysine sur position 6 du cycle purique du dérivé thiolé²²¹.

L'analogie structurale des dérivés thiolés avec les bases auxquelles ils correspondent leur permet de s'apparier avec les bases complémentaires *via* les liaisons hydrogènes classiques de Watson-Crick. Des études cristallographiques ont montré que la longueur de la liaison $\text{C}_4 = \text{S}$ et le rayon de Van der Waals de l'atome de soufre sont plus importants que ceux observés pour la liaison $\text{C}_4 = \text{O}$ de la base non thiolée²²². Le remplacement de la thymidine par son analogue thiolé semble n'avoir qu'un effet limité sur la stabilité des oligonucléotides lorsque le taux de substitution est faible²²³. La substitution de la désoxyguanosine a des effets déstabilisateurs plus prononcés et influe sur la structure des duplex formés²²⁴. De plus, un taux élevé de substitution par la ^{45}T altère fortement les propriétés de l'oligonucléotide²²⁵.

Concernant les avantages des dérivés soufrés dans le cadre du photomarquage d'affinité, on peut citer l'analogie structurale, qui permet de les incorporer dans l'ADN sans perturbations majeures des interactions à étudier. De plus, cela permet d'assurer le photomarquage de résidus directement en contact avec l'acide nucléique : on dit qu'il s'agit de sondes « zero-length ». Les oligonucléotides contenant la ^{45}T sont stables pendant des années s'ils sont conservés à l'abri de la lumière et à -80°C . L'irradiation a lieu à des longueurs d'ondes supérieures à 300 nm, ce qui évite d'endommager les protéines et les acides nucléiques. Enfin, la durée de vie de leur état excité est inférieure à la microseconde et il a une forte capacité à former des adduits covalents avec des acides aminés. Toutes ces raisons font des analogues thiolés d'acides nucléiques et plus particulièrement de la ^{45}T une bonne sonde photoactivable.

Parmi les inconvénients, on observe expérimentalement que la ^{45}T est sensible à l'introduction d'un isotope radioactif à des fins d'analyse. On remarque en effet que le rendement de photomarquage des sondes contenant la ^{45}T marquées au phosphore radioactif

diminue de manière continue après le marquage radioactif pour devenir quasiment nul après 3 semaines, ce qui est problématique si on veut utiliser la radioactivité comme méthode de détection.

1.3.2 Dérivés halogénés de l'uridine

La 5-Bromouridine ($^{5\text{Br}}\text{U}$), la 5-Iodouridine ($^{5\text{I}}\text{U}$), la 5-Bromo-2'-desoxyuridine ($^{5\text{Br}}\text{dU}$) et la 5-Iodo-2'-desoxyuridine ($^{5\text{I}}\text{dU}$) sont des analogues de l'uridine utilisés pour le photomarquage d'affinité (Figure 97). L'irradiation à des longueurs d'ondes comprises entre 300 et 325 nm leur permet de se lier de façon covalente avec les protéines présentes à proximité. Les dérivés bromés de l'uridine sont utilisés depuis les années 70 et ont ensuite été remplacés par les dérivés iodés, plus efficaces, à partir des années 1990. Ainsi, l'utilisation d'une séquence d'ADN contenant l'opéron lactose où les thymidines étaient remplacées par des $^{5\text{Br}}\text{U}$ a ainsi permis l'identification de points de contacts avec le répresseur de l'opéron lactose²²⁶. La 5IU a également été utilisée pour identifier le site d'interaction entre une séquence d'ARN en « épingle à cheveux » et le domaine *N*-terminal de la protéine U1A²²⁷.

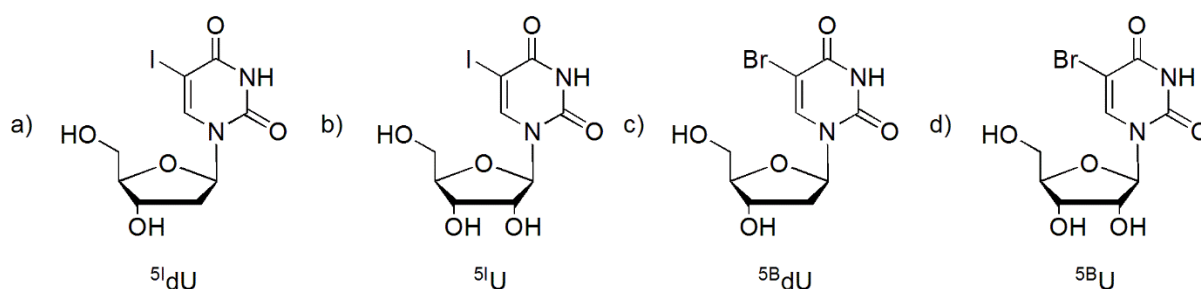


Figure 97. Dérivés halogénés de l'uridine ; a) 5-Iodo-2'-désoxyuridine b) 5-Iodouridine c) 5-Bromo-2'-désoxyuridine d) 5-Bromouridine.

Deux mécanismes différents sont à l'origine de la formation d'adduits covalent selon la nature de l'halogène en position 5. Dans le cas de la $^{5\text{Br}}\text{U}$, l'irradiation à 308 nm donne lieu à un état singulet qui va se convertir en état triplet par croisement intersystème. C'est cet état triplet qui va réagir par des mécanismes de transfert d'électrons²²⁸. Un autre mécanisme possible, mis en évidence dans une étude avec un analogue de la Tyrosine, met en jeu l'excitation de la tyrosine à une longueur d'onde de 325 nm puis la réaction avec l'uridine, plutôt que le mécanisme inverse²²⁸. Pour la $^{5\text{I}}\text{U}$, c'est un état singulet qui va être responsable de la formation de la liaison covalente. L'absorption d'un photon va entraîner la rupture homolytique de la liaison C-I et la formation de liaisons covalentes avec les protéines

présentes à proximité²²⁹. La comparaison entre l'efficacité de photomarquage de la ^{51}U et la $^{5\text{B}}\text{U}$ indique que la ^{51}U est entre deux et cinq fois plus efficace que la seconde^{227,230}.

Le photomarquage de nombreux acides aminés a été rapporté : Tryptophane, Histidine, Tyrosine, Phénylalanine et Alanine pour les dérivés bromés; Tyrosine, Histidine et Phénylalanine pour les dérivés iodés. Il est à noter que d'autres mécanismes que ceux observés en solution peuvent intervenir dans le photomarquage de protéines : pour la ^{51}dU , il a été observé des rendements de photomarquage de 90% alors que les études en solutions indiquent un rendement maximal de 66%. L'explication fait intervenir des interactions de type empilement π favorisées par la structure de la protéine qui augmenterait l'efficacité de photomarquage²²⁹.

L'analogie structurale de la ^{51}dU et de la $^{5\text{Br}}\text{U}$ avec la thymidine sont un des avantages de ces groupements photoactivables. En effet, les rayons de Van der Waals des atomes de brome et d'iode (respectivement 1.95 Å et 2.15 Å) sont proches de celui du groupement méthyle (2.0 Å)²³¹. De manière similaire à la $^{4\text{S}}\text{T}$, cela permet de ne pas perturber l'interaction entre les sondes photoactivables et les protéines d'intérêt. Un des désavantages sont les longueurs d'ondes spécifiques nécessaires à l'activation des dérivés halogénés. L'excitation à 308 ou 325 nm nécessite souvent l'utilisation de lasers qui ne sont pas des matériels accessibles facilement.

Nous avons donc vu qu'il existe de nombreux groupements photoactivables, certains spécifiques à l'étude des interactions mettant en jeu les acides nucléiques. Chacun possède des avantages et des inconvénients qui lui sont propres dans le cadre du photomarquage d'affinité. Le chapitre suivant sera consacré aux différentes méthodes de synthèse de sondes oligonucléotiques photoactivables.

1.4 Introduction des groupements photoactivables dans des sondes oligonucléotiques

L'introduction de la fonction photoactivable dans une séquence d'ADN cible peut se faire de différentes manières, par voie enzymatique *via* l'utilisation d'enzymes impliquées dans la synthèse des acides nucléiques (polymérases, terminal transférases, ligases), ou par voie chimique *via* la synthèse automatique d'oligonucléotides modifiés.

1.4.1 Synthèse enzymatique de sondes oligonucléotidiques photoactivables

Les dérivés triphosphates des analogues nucléosidiques photoactivables peuvent être incorporés dans des fragments d'ADN ou d'ARN par des méthodes enzymatiques. Les polymérases de l'ADN peuvent ainsi utiliser des analogues photoactivables dans leur processus d'élongation de l'ADN. Lorsque les bases photoactivables ont une forte analogie structurale avec les bases non modifiées, comme les bases décrites en 1.3.1 et 1.3.2, cette incorporation se fait sans problème. Dans le cas de la ^{45}T , une activité plus efficace (comparée à la thymidine) d'incorporation par le fragment de Klenow de la polymérase I d'*E.coli* a même été rapportée²²³. L'utilisation de mélanges entre nucléotides triphosphates modifiés ou naturels permet une introduction plus ou moins grande de l'analogue photoactivable dans la séquence d'ADN, mais sans pouvoir contrôler précisément la position des modifications dans la séquence cible sauf si l'on atteint le taux de substitution maximal.

Si l'on souhaite introduire le nucléoside photoactivable à une position déterminée dans la séquence de la sonde, deux méthodes sont utilisables selon la position désirée de l'analogue nucléotidique. La plus simple consiste à introduire l'analogue à l'extrémité 3' de la séquence d'ADN cible, à l'aide du nucléotide triphosphate modifié et d'une ADN polymérase.

L'introduction du nucléotide modifié à l'intérieur de la séquence à une position spécifique est plus compliquée. L'équipe de Lavrik a par exemple utilisé cette méthodologie dans de nombreux travaux concernant l'étude de complexes ADN-protéines et notamment des systèmes de réparation²³² via l'utilisation de (d)NTPs photoactivables²³³. Le principe de l'incorporation d'un dNTP au milieu d'une séquence est présenté dans la figure 98²³⁴. Deux brins d'ADN, P1 et P2, correspondant chacun à une partie de la séquence cible sont maintenus à proximité par un petit brin complémentaire qui va servir de « pont » et de matrice pour incorporer le nucléotide modifié. Le nucléotide photoactivable est alors incorporé par la polymérase β à l'extrémité 3' du brin P1. Les deux brins d'ADN sont ensuite liés par l'action de l'ADN ligase du phage T4, formant la séquence d'ADN cible. Le duplex est dénaturé et le brin contenant le nucléotide modifié est purifié par électrophorèse.

Figure 98. Principe de l'introduction d'une sonde photoactivable à une position spécifique par voie enzymatique. D'après ²³⁴.

Cette méthode comporte de gros désavantages : l'utilisation d'enzymes successives pour la biosynthèse des sondes est contraignante et implique des réactions sur de faibles quantités d'ADN. Par ailleurs, cette méthode est limitée par la haute spécificité des enzymes qui peuvent avoir des difficultés à reconnaître le nucléotide modifié.

1.4.2 Synthèse chimique de sondes oligonucléotidiques photoactivables

Compte tenu des spécificités de la synthèse chimique de l'ADN, deux voies sont possibles pour l'introduction de fonctions photoactivables dans des séquences d'ADN. On peut soit utiliser des synthons phosphoramidites portant déjà des fonctions photoactivables (stratégie de fonctionnalisation pré-synthétique), ou bien greffer les fonctions photoactivables en fin de synthèse automatique sur l'oligonucléotide (stratégie de fonctionnalisation post-synthétique) (Figure 99).

Figure 99. Les deux stratégies principales de fonctionnalisation d'oligonucléotides : A) Marquage pré-synthétique ; B) Marquage post-synthétique. (Les nucléotides sont représentés en bleu ; les fonctions réactives sont représentées sous forme de bâtonnets rouges et les modifications introduites sont représentées sous forme de ronds rouges). D'après ²³⁵.

Le principe de ces deux approches sera illustré en s'appuyant sur des exemples de la littérature.

1.4.3 Stratégie pré-synthétique : introduction du groupement photoactivable via les synthons phosphoramidites

La modification pré-synthétique des oligonucléotides consiste à introduire la modification souhaitée (ici, l'agent de photomarquage) sur le synthon phosphoramidite. Ce

dernier est ensuite utilisé dans la synthèse sur support solide et incorporé à la position souhaitée. Le pré-requis pour l'utilisation de cette stratégie est que le phosphoramidite devra être suffisamment stable pour résister aux conditions de synthèse et de déprotection (milieu acide, basique, conditions oxydantes) des oligonucléotides. Dans le paragraphe qui suit, nous allons présenter quelques exemples de cette stratégie appliqués à la préparation de sondes d'acides nucléiques photoactivables.

1.4.3.1 Introduction directe de nucléobases photoactivables

En ce qui concerne l'incorporation des bases photoactivables mentionnées dans le paragraphe 1.3, il est possible d'accéder aisément aux synthons phosphoramidites correspondants, leur synthèse étant bien maîtrisée^{236,237}. Ces synthons sont d'ailleurs accessibles commercialement, ce qui rend la méthode phosphoramidite incontournable si l'on souhaite incorporer ces bases photoactivables par la voie chimique. La Figure 100 présente plus particulièrement deux phosphoramidites disponibles chez Glen Research. Il est à noter que le groupement protecteur cyanoéthyl du phosphoramidite pour l'incorporation de la ⁴S^T nécessite une étape de déprotection supplémentaire au DBU.

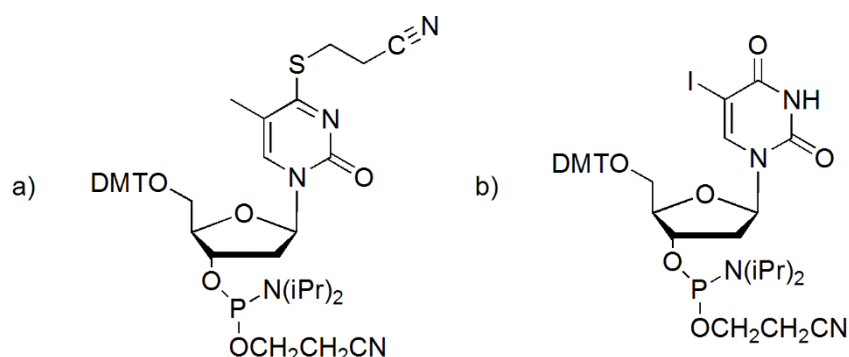


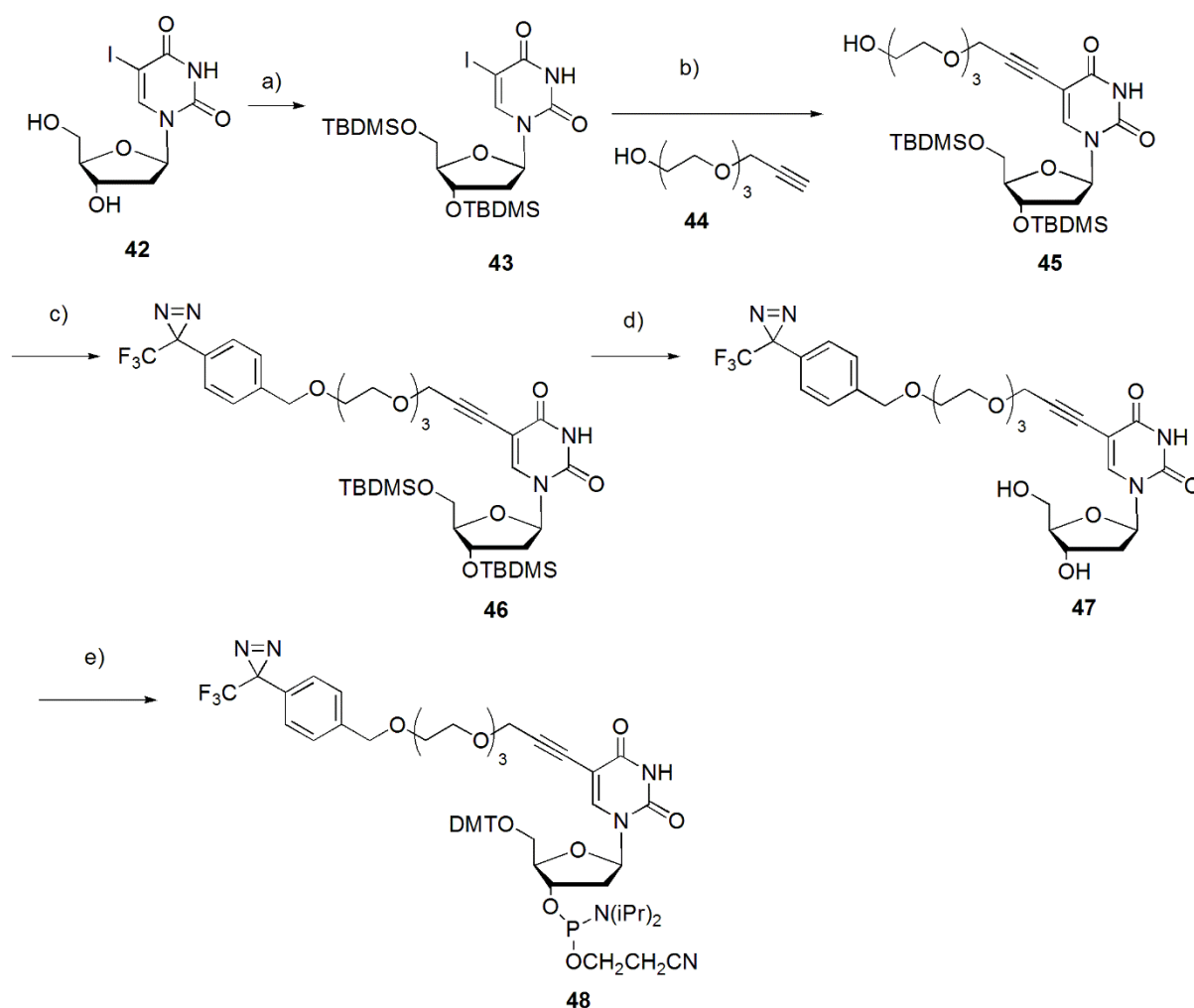
Figure 100. Phosphoramidites disponibles commercialement (Glen Research) ; a) phosphoramidite pour l'incorporation de la ⁴S^T ; b) phosphoramidite pour l'incorporation de la ⁵I^dU.

1.4.3.2 Incorporation d'agent de photomarquage *via* des bras espaceurs sur le nucléoside

Dès qu'on souhaite introduire d'autres groupements photoactivables via des bras espaceurs sur la partie sucre ou base du nucléoside, la synthèse de phosphoramidites plus

complexes doit être envisagée avec une fonctionnalisation sur la partie sucre ou la partie base nucléique qui nécessite la mise en œuvre d'un jeu de protection/déprotection important.

La figure 101 présente un exemple de synthèse d'un phosphoramidite permettant l'introduction d'une fonction trifluorométhylphényldiazirine dans un analogue de la thymidine sur la position 5 de la base, réalisée par le groupe de Thomas Carell²³⁸. La synthèse commence par la protection des alcools de la ⁵¹dU par des groupements TBDMS, suivie d'un couplage de Sonogashira qui permet l'introduction d'un espaceur triéthylène glycol auquel on vient coupler ensuite la trifluorométhylphényldiazirine pour donner le composé **46**. Les étapes finales sont la déprotection des alcools, la protection de l'alcool primaire en 5' par un groupement DMT et la phosphitylation finale. Une fois synthétisé et purifié, le phosphoramidite **48** peut être utilisé lors d'un cycle de synthèse automatique pour être introduit dans une séquence oligonucléotidique. Après complétion de la synthèse, clivage du support et purification par HPLC, la sonde oligonucléotidique peut être utilisée. Dans cet exemple, cette sonde photoactivable a été complexée à un brin complémentaire comportant une 8-Oxo-2'-désoxyguanosine, un dommage oxydatif de l'ADN, ce qui a permis d'identifier de nouvelles protéines interagissant avec ce dommage oxydatif de l'ADN.

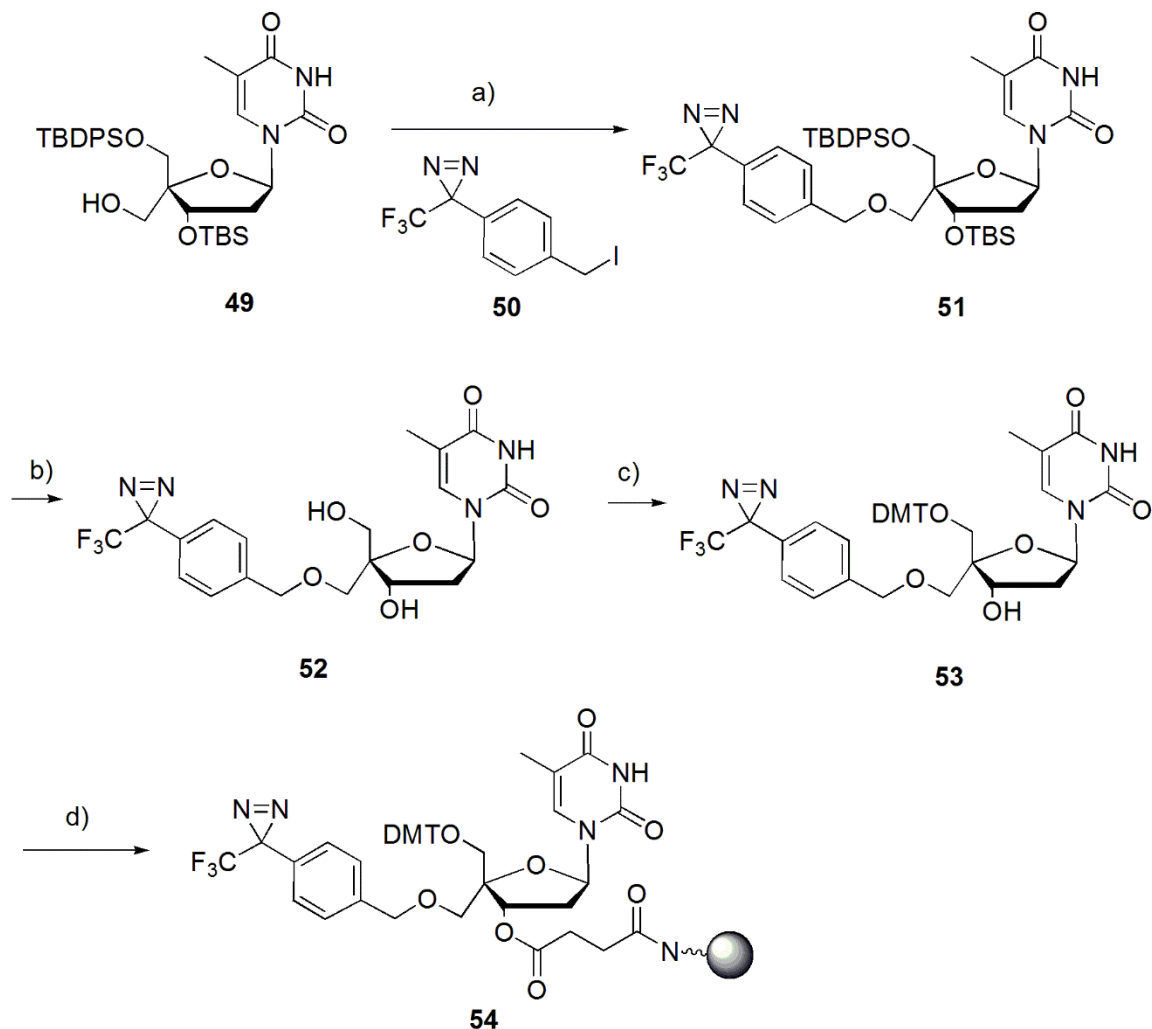


a) TMDMSCl, DMAP, rt, 4h, 95%; b) **44**, Pd(Ph₃)₂Cl₂, CuI, DMF, iPrNEt₃, RT, 12h, 64%; c) NaH, THF, rt, 3.5h, 67%; d) TBAF, THF, rt, 2h, 100%; e) i) DMTCl, pyridine, rt->0°C, 2h, 40% ii) (iPr)₂N₂-P-O-CH₂CH₂CN, diisopropylammonium tetrazolate, CH₂Cl₂, RT, 3h, 55%

Figure 101. Synthèse du synthon phosphoramidite d'un analogue de thymidine comportant un groupement photoactivable en position 5 du cycle pyrimidique²³⁸.

Pour illustrer l'introduction d'un groupement photoactivable sur la partie sucre d'un nucléoside, on peut citer les travaux de Lebman et al. (Figure 102)²³⁹. Le réactif **49** est obtenu en quelques étapes de synthèse à partir de la 5'-formylthymidine²⁴⁰. L'alcool introduit en position 4' est sélectivement fonctionnalisé par une trifluorométhylphényl diazirine à partir du dérivé iodé **50** correspondant. Les alcools en position 5' et 3' du composé **51** sont ensuite déprotégés avec du TBAF. L'alcool primaire en position 5' de **52** est ensuite protégé avec un groupement DMT, puis le synthon est greffé sur un support solide avec des agents de couplage. Les auteurs ont ensuite utilisé cette résine **54** dans des cycles de synthèse automatique pour obtenir des sondes oligonucléotidiques comportant un analogue de thymidine photoactivable à leur extrémité 3'. L'intérêt de l'introduction à la position 4' du

sucres est qu'elle permet de projeter le groupement photoactivable dans le petit sillon de l'ADN à proximité des liaisons phosphodiester, qui est une région ciblée par de nombreuses enzymes de réplication de l'ADN. Les auteurs ont ensuite utilisé leur sonde pour photomarquer l'ADN polymérase β humaine.



a) **50**, K_2CO_3 , DMF, 73%; b) TBAF 1M, THF, 77%; c) DMTCl, DMAP, pyridine, 47%; d) succinyl-LCAA-CPG, DMAP, EDC, pyridine, 4-nitrophenol, triethylamine.

Figure 102. Synthèse et greffage sur colonne d'un analogue de thymidine comportant un groupement photoactivable en position 4' du sucre (LCAA-CPG : Long Chain Aminoalkyl-Controlled Pore Glass).

L'introduction de groupements photoactivables *via* la stratégie pré-synthétique présente certains aspects contraignants. Il faut trouver des groupements protecteurs adaptés à la fonction photoactivable que l'on souhaite introduire, ou bien que cette fonction photoactivable soit stable lors des étapes de la synthèse du phosphoramidite ainsi que pendant la synthèse automatique qui a lieu dans des conditions assez drastiques. De plus, si l'on

souhaite varier la nature de la sonde photoactivable ou de l'éventuel bras espaceur, il faut resynthétiser le phosphoramidite correspondant, ce qui présente à chaque fois un effort de synthèse non négligeable. Par ailleurs, un problème majeur qui peut survenir lorsque des phosphoramidites hautement modifiés sont utilisés dans un cycle de synthèse automatique sur support solide, est la chute de rendement de synthèse lors du couplage de ce synthon qui peut diminuer drastiquement le rendement en oligonucléotide final.

1.4.4 Introduction du groupement photoactivable par la fonctionnalisation post-synthétique

Pour remédier à ces inconvénients, on peut choisir d'introduire le groupement photoactivable *via* des réactions de fonctionnalisation post-synthétiques. Cette voie repose sur l'introduction pendant la synthèse automatique d'analogues nucléosidiques ou de blocs phosphoramidites comportant un petit groupement fonctionnel qui permettra ensuite l'introduction de molécules d'intérêt (ici, l'agent de photomarquage) après la fin des cycles de synthèse de la séquence oligonucléotidique.

La fonctionnalisation post-synthétique implique donc l'introduction dans la séquence nucléique de nucléosides portant des fonctions, masquées ou non, qui vont pouvoir réagir de manière orthogonale au reste des fonctions présentes dans l'oligonucléotide et permettre l'introduction de groupements spécifiques.

De nombreuses méthodes de marquage covalent d'oligonucléotides ont été décrites (pour revue, voir ²⁴¹). Certaines impliquent la fonctionnalisation avant le clivage de l'oligonucléotide du support solide sur lequel il est attaché, et les réactions peuvent alors avoir lieu sur l'oligonucléotide encore protégé. C'est le cas de la technique dite des nucléosides convertibles qui a été introduite par le groupe de Verdine²⁴². Il est également possible de fonctionnaliser les oligonucléotides après clivage du support solide. La méthode la plus ancienne consiste à introduire une fonction amine ou thiol à l'extrémité 5' qui pourra réagir ensuite avec des esters activés, des maléimides, des fonctions disulfures ou des dérivés halogénés.

Plus récemment, avec le développement des réactions de conjugaison bioorthogonales, des techniques de marquage utilisant les réactions de Diels Alder, de Huisgen ou la ligation de

Staudinger ont été développées. Dans ce paragraphe, nous présenterons quelques exemples de modifications post-synthétiques appliqués à la préparation de sondes photoactivables.

1.4.4.1 Méthode du nucléoside convertible

La méthode dite du « nucléoside convertible » est une des méthodes de fonctionnalisation post-synthétiques de l'ADN les plus employées. Introduite et développée par le groupe de Verdine²⁴², elle consiste en l'introduction d'analogues nucléosidiques contenant un groupement partant qui va pouvoir être déplacé par un nucléophile après la synthèse automatique. Le choix du nucléophile permet d'introduire la fonction désirée sur la séquence d'ADN cible. Les nucléosides peuvent être introduits à la position 4 des cycles pyrimidiques ou aux positions 2 et 6 des purines, ce qui permet de conserver les interactions hydrogène nécessaires à la stabilisation des doubles brins d'ADN, comme illustré dans la Figure 103²⁴³. Les différentes positions modifiables permettent aussi de projeter le groupement soit dans le grand sillon de l'ADN (N⁶-alkyl-A, N⁴-alkyl-C) ou dans le petit sillon (N²-alkyl-G).

Figure 103. Projection des fonctionnalisations dans le grand ou le petit sillon de l'ADN en fonction de la position et de la base fonctionnalisée. D'après²⁴³.

Cette méthode peut s'appliquer tant à la fonctionnalisation de séquence d'ADN que d'ARN. Un des premiers exemples d'utilisation de cette méthode a été la synthèse de duplex contraints stériquement pour l'étude de la DNMT bactérienne M.HaeIII²⁴⁴. Deux brins complémentaires ont été chacun fonctionnalisés par une diamine contenant un pont disulfure (Figure 104). Après la fonctionnalisation post-synthétique, les deux brins ont été réduits avec du DTT, les deux brins complémentaires ont été appariés puis oxydés. L'oxydation a permis de lier les deux brins complémentaires par un pont disulfure et de former les analogues contraints, qui ont ensuite été évalués vis-à-vis de l'enzyme bactérienne.

Figure 104. Synthèse d'analogues contraints de duplex de l'ADN ; A) Fonctionnalisation post-synthétique de brins d'ADN comportant la 2-Fluoro-2'-désoxyinosine ; B) Structure des sondes et des contrôles utilisées pour étudier l'enzyme M.HaeIII ; Tiré de ²⁴⁴.

De nombreux analogues de nucléosides ont été synthétisés pour être utilisés dans le cadre de cette méthode de fonctionnalisation post-synthétique. Ils sont présentés dans la Figure 105. Tous comportent des bons groupes partants qui vont pouvoir être déplacés par des amines primaires. Les conditions typiques de fonctionnalisation impliquent des concentrations élevées du nucléophile, un solvant comme le DMSO ou le DMF et des durées de réactions élevées. Après fonctionnalisation de la séquence par le nucléophile désiré, les groupements protecteurs spécifiques sont déprotégés si nécessaire puis l'oligonucléotide est clivé de la colonne et purifié par HPLC. Dans certains cas, la fonctionnalisation, déprotection et clivage de la colonne peuvent s'effectuer en une seule étape²⁴⁵, simplifiant grandement la procédure de synthèse.

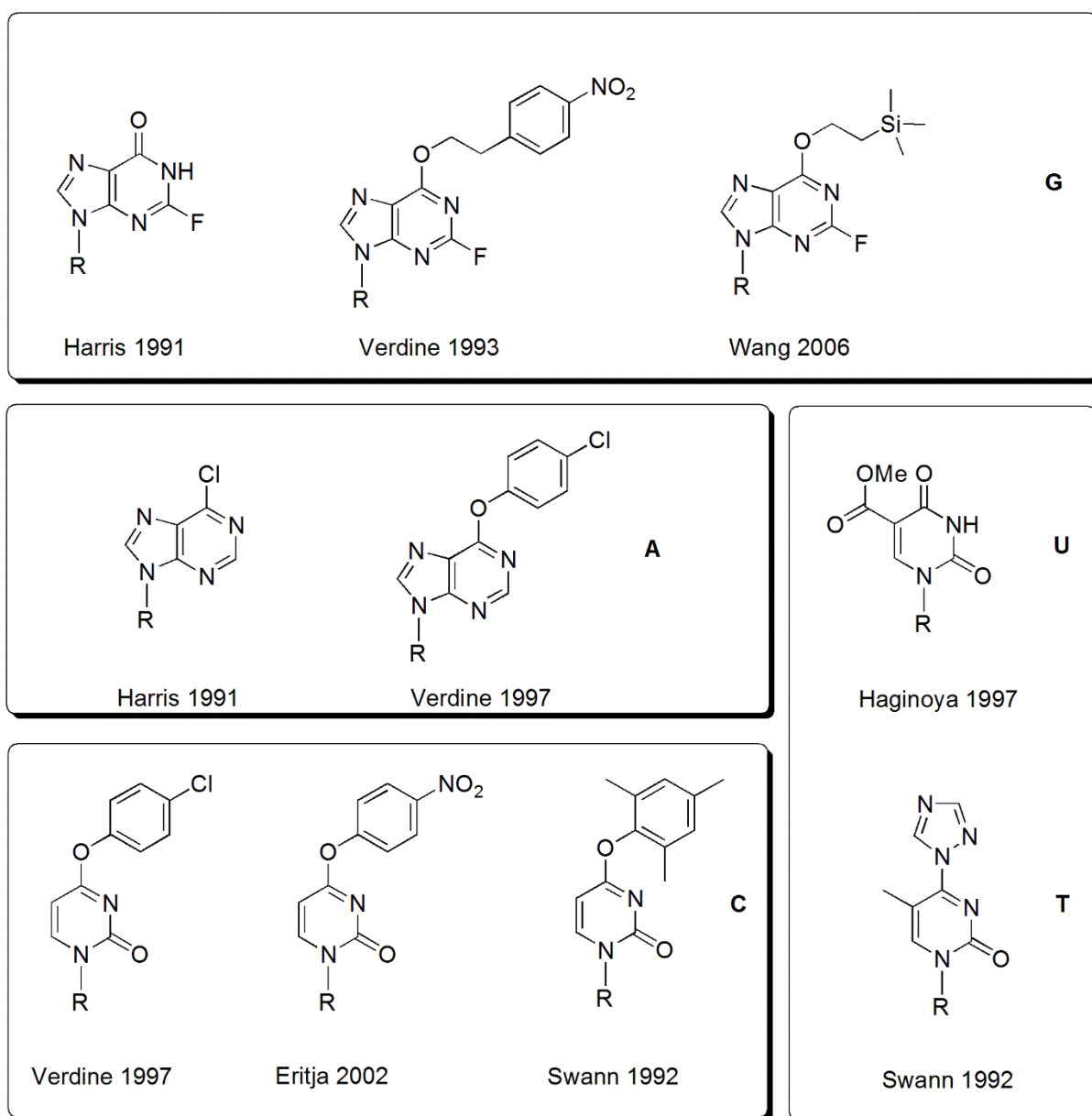


Figure 105. Différents nucléosides convertibles utilisés dans le cadre de la méthode dite du « convertible nucleoside » ; Analogues de G^{244,246,247} ; Analogues de A^{246,248} ; Analogues de C²⁴⁸⁻²⁵⁰ ; Analogues de U²⁵¹ ; Analogues de T²⁵⁰.

Dans le cadre du photomarquage d'affinité, la méthode du nucléoside convertible a par exemple été utilisée pour introduire des diazirines aliphatiques dans le petit et le grand sillon de duplex de l'ADN (Figure 106). Par cette méthode, la fonctionnalisation post-synthétique de séquences contenant la 2-fluoro-2'-désoxyinosine ou la O⁶-triazolyl-2'-désoxyuridine a permis de photomarquer la N-adénine méthyltransférase d'*E. coli* (EcoDam) et la O⁶-alkylguanine-ADN alkyltransférase (hAGT) hAGT²⁵².

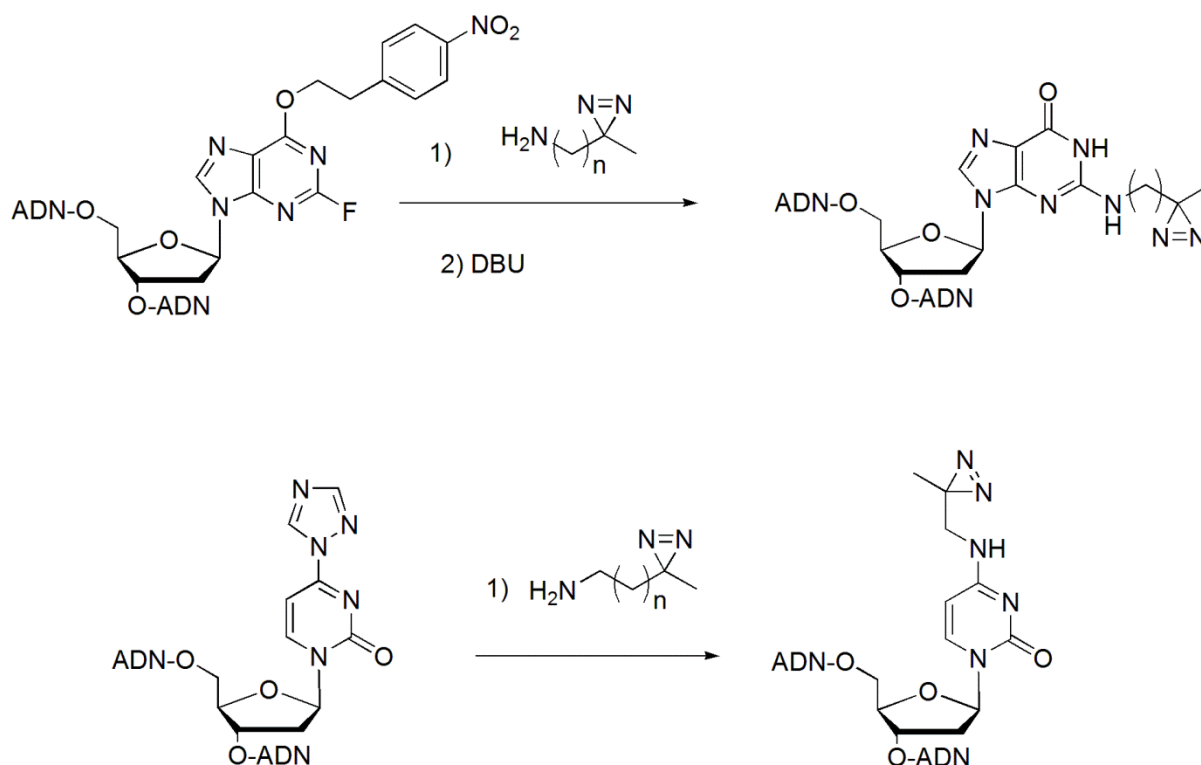


Figure 106. Fonctionnalisation post-synthétique de la 2-Fluoro-2'-désoxyinosine et de l'O⁶-triazolyl-2'-désoxyuridine par des diazirines alyphatiques.

Un autre exemple d'application des nucléosides convertibles concerne l'introduction de la ⁴⁵T. Même si l'introduction de celle-ci par voie chimique est effectuée quasi exclusivement par le moyen de phosphoramidites, il est possible de l'obtenir *via* une réaction de modification post-synthétique (Figure 107). La 4-Triazolo-thymidine peut ainsi être convertie en 4-Thiothymidine par l'action de l'acide thioacétique à température ambiante suivie d'une déprotection au DBU dans le méthanol²⁵⁰. Il est aussi possible d'introduire des alcools ou des hydrazynes à cette position 4 du cycle pyrimidique.

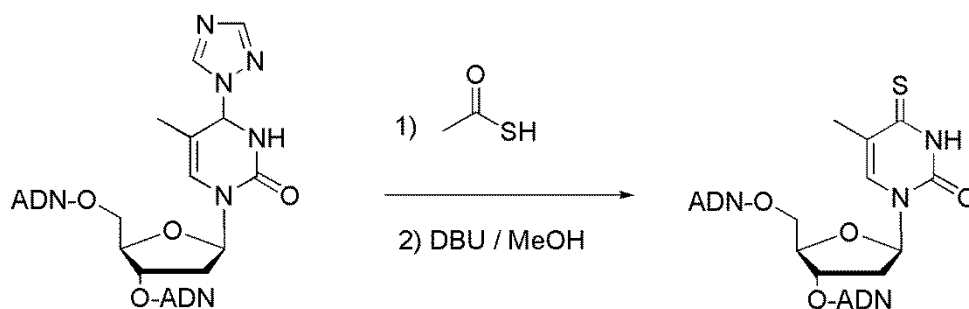


Figure 107. Conversion post-synthétique de la 4-Triazolo-thymidine en 4-Thiothymidine.

Enfin, notons que la 4-thio-2'-désoxyuridine peut, en milieu faiblement basique, réagir comme un nucléophile vis-à-vis de dérivés halogénés ou d'électrophiles (Figure 108)^{253,254}. L'exploitation de cette réaction a permis de ponter deux brins complémentaires dans un duplex²⁵⁵ mais, à notre connaissance, aucun groupement photoactivable n'a été introduit par cette méthode.

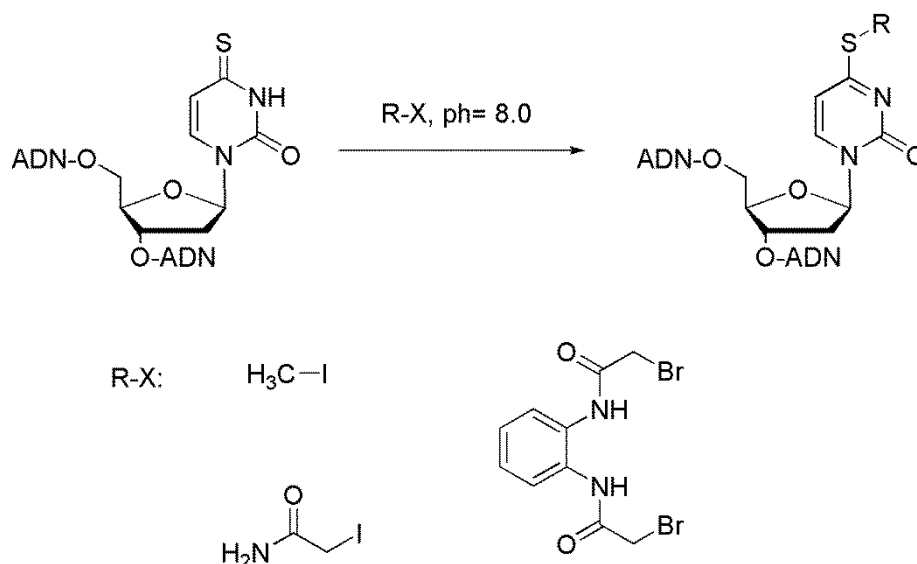


Figure 108. Fonctionnalisation post-synthétique de la ^{4S}dU en milieu basique par des dérivés halogénés.

1.4.4.2 Réactions bioorthogonales

La 5-Iodo-2'-désoxyuridine peut par exemple être fonctionnalisée par une réaction de couplage de Suzuki-Miyaura dans des conditions de réaction non destructives pour l'ADN, en utilisant des dérivés d'acide boronique et une catalyse au palladium²⁵⁶. Cette réaction a permis d'introduire sur la position 5 du cycle pyrimidique des groupements photoactivables comme des benzophénones ou des diazirines, ainsi que d'autres fonctionnalités comme des sucres ou des pyrènes (Figure 109). Nous reviendrons sur ces travaux dans la conclusion de ce chapitre.

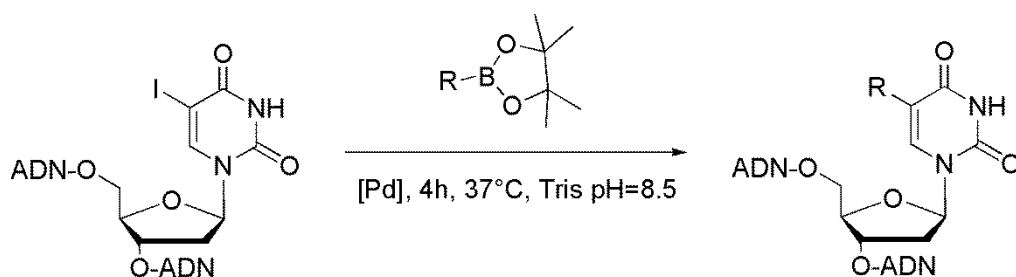


Figure 109. Fonctionnalisation post-synthétique de la 5-Iodo-2'-desoxyuridine par une réaction de couplage catalysée au palladium.

Les cycloadditions sont aussi des réactions de choix pour fonctionnaliser les oligonucléotides. La 7-Vinyl-7-deazaguanosine peut ainsi être modifiée de façon post synthétique par une réaction de Diels-Alder avec des maléimides (Figure 110). Cette cyclisation a permis par exemple d'introduire des groupements de type benzophénone, TEMPO, pyrène ou biotine pour des applications de photomarquage d'affinité, de marquage de spin ou de fluorescence²⁵⁷.

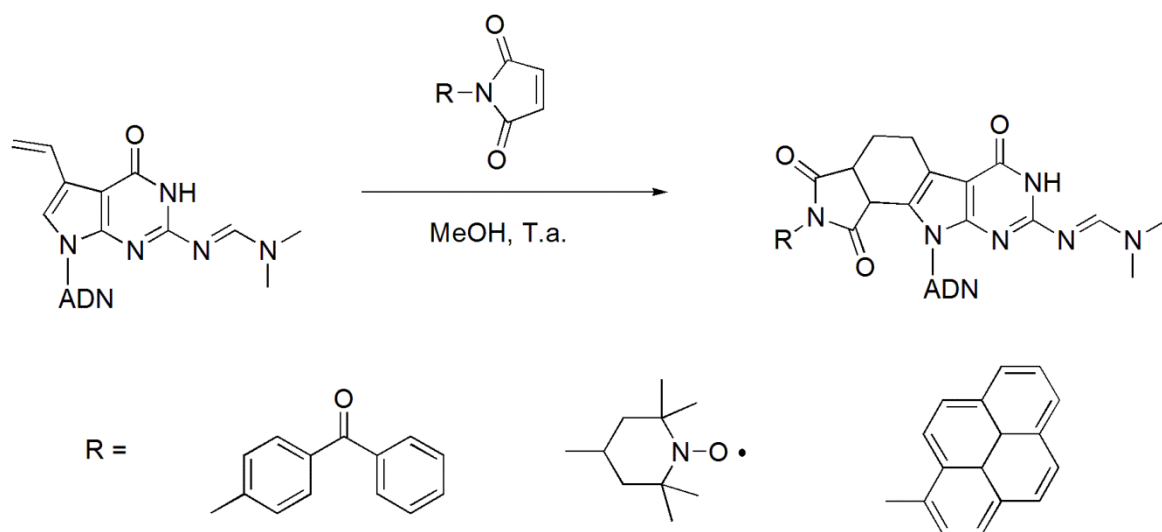


Figure 110. Fonctionnalisation post-synthétique de la 7-Vinyl-7-deazaguanosine par une réaction de Diels-Alder.

Les différentes méthodes de fonctionnalisation post-synthétiques permettent donc l'introduction relativement simple d'une grande diversité de groupements dans des séquences d'ADN. De nombreux nucléosides modifiés sont disponibles commercialement dans cette

optique, et s'ajoutent à ceux dont la synthèse et l'utilisation a été décrite. Cette versatilité permet de préparer rapidement des sondes pour des applications différentes, ou de moduler la projection dans l'espace des groupements désirés sans avoir à synthétiser de nombreux phosphoramidites. La fonctionnalisation post-synthétique est donc une méthode de choix pour l'introduction de groupements photoactivables dans des sondes oligonucléotidiques.

2 Conception, synthèse et évaluation de sondes photoactivables pour l'identification de protéines partenaires de la 5hmC

2.1 Conception des sondes

L'objectif de cette étude vise à mettre en place de nouveaux outils pour l'identification dans des extraits cellulaires de protéines interagissant avec la 5-hydroxyméthylcytosine dans un modèle oligonucléotidique. Le principe de nos sondes repose sur l'introduction de modifications sur les deux brins complémentaires du duplex. L'un des brins du duplex va porter le « modulateur épigénétique », la 5-hydroxyméthylcytosine. Cette modification sera responsable de l'interaction observée entre les sondes et les protéines présentes. Le brin complémentaire est lui modifié de façon à contenir d'une part, un groupement photoactivable à proximité de la modification épigénétique, d'autre part, un groupement biotine à son extrémité 3'. Le groupement photoactivable va permettre la création de liaisons covalentes avec les protéines interagissant préférentiellement avec la 5-hydroxyméthylcytosine ; la biotine permettra d'enrichir et de purifier ces adduits covalents protéines-sonde à partir des extraits cellulaires (Figure 111).

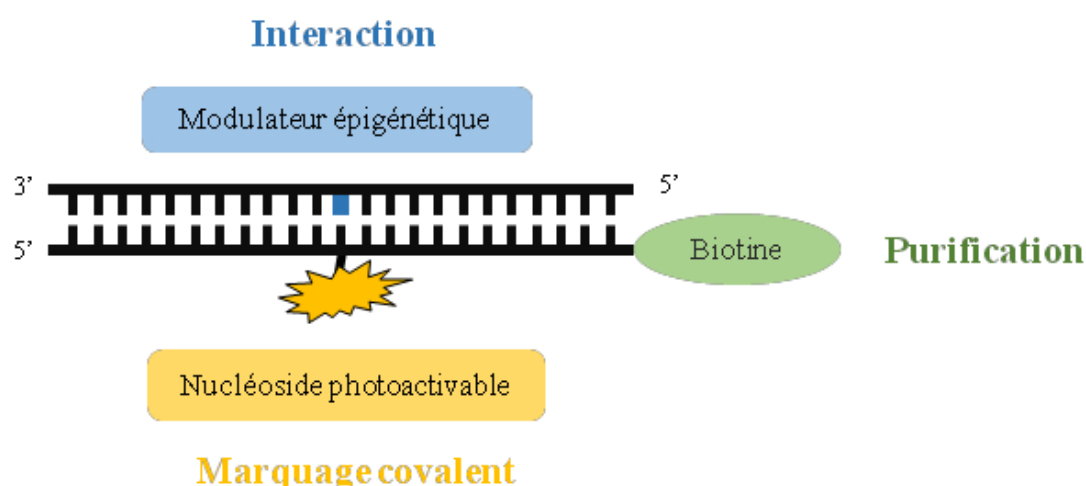


Figure 111. Le principe des sondes photoactivables et le rôle des trois modifications introduites.

Comme l'indique la figure 110, trois modifications doivent donc être introduites dans nos duplex que nous avons conçus de la façon suivante.

- Choix de la séquence comportant la 5hmdC :

Lorsque nous avons commencé ces travaux, peu d'éléments étaient connus concernant les séquences privilégiées contenant la 5hmC. Comme les 5hmC sont issues de l'oxydation des 5mC, il nous a paru judicieux d'utiliser une séquence comportant un site CpG reconnue par les ADN méthyltransférases. Nous avons donc choisi d'incorporer la 5hmC dans la séquence modèle étudiée précédemment au laboratoire dans le cadre de l'étude sur la méthylation de l'ADN. Afin de permettre aux protéines d'avoir un site de fixation sur notre duplex, nous avons choisi de préparer des séquences suffisamment longues, 30-mères et de placer le site 5hmCpG au milieu. Les contrôles comportant la base C et 5mC ont également été prévus (Figure 112).

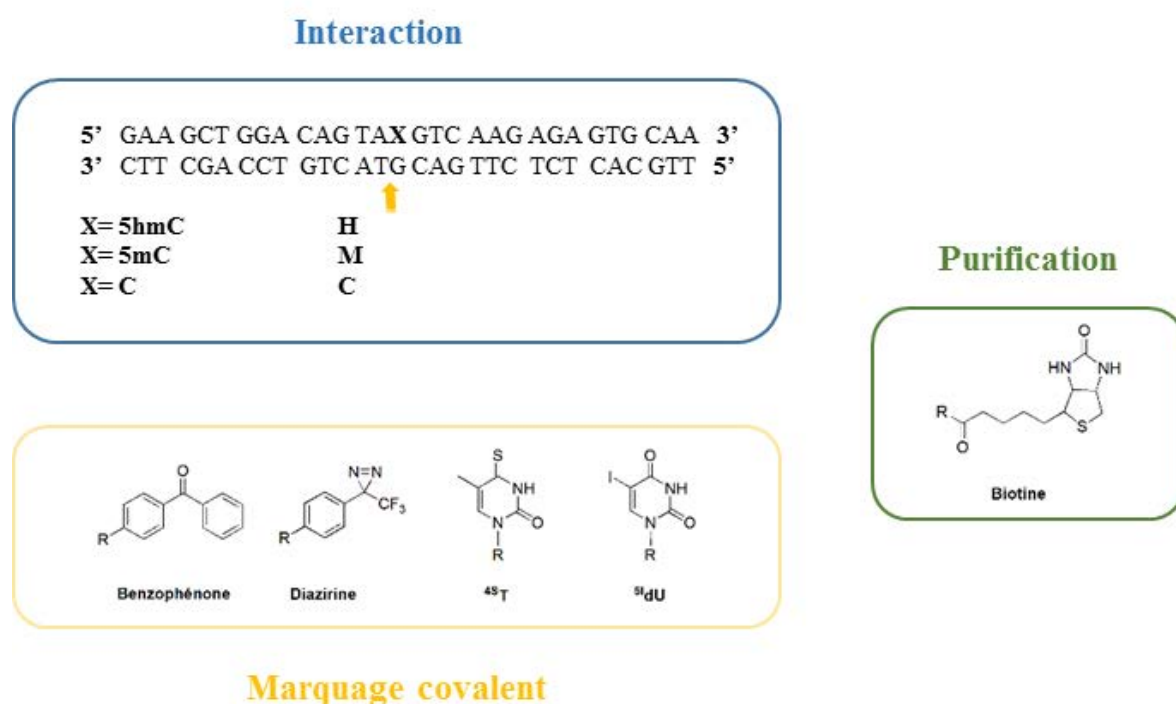
- Choix des agents photoactivables :

Afin de disposer d'une panoplie de différents agents photoactivables (qui seraient susceptibles de donner lieu à des résultats différents), nous nous sommes proposé d'introduire différentes molécules dans notre brin complémentaire et selon deux approches différentes : comme nous l'avons vu dans le paragraphe **1.2**, la trifluorométhylphényldiazirine et la benzophénone représentent les meilleures fonctions photoactivables classiques. Afin de les incorporer, nous avons fait le choix d'utiliser la méthode des nucléosides convertibles de manière post-synthétique évoquée dans le paragraphe **1.4.4.1**. Cette méthode nous paraissait adaptée pour moduler facilement la nature des groupements photoactivables introduits et nous les introduirons *via* un bras espaceur. Nous souhaitons également disposer d'un oligonucléotide moins modifié et donc utiliser des nucléosides photoactivables faiblement fonctionnalisés : notre choix s'est porté sur la 5-iodo-2'-désoxyuridine et la 4-thiothymidine, notre travail étant mené en collaboration avec le Dr. Carole Saintomé, très familière avec l'utilisation de ce dernier nucléoside.

- Choix du point d'ancrage de la biotine :

Le brin comportant la sonde photoactivable sera celui sur lequel les protéines seront fixées par une liaison covalente. Du fait du rendement parfois faible de la réaction de

photomarquage ou de la faible affinité de certains partenaires protéiques pour l'ADN, un enrichissement d'affinité de type biotine/streptavidine sera indispensable pour identifier les protéines liées à la sonde. La biotine sera fixée à l'extrémité 3' de notre oligonucléotide *via* un bras espaceur lors de la synthèse automatique sur support solide.



glycol (PEG) commercialisée par Glen Research (Figure 113). La position N¹ du cycle uréido de la biotine est protégée avec un groupe *t*-butylbenzoyl, ce qui permet d'éviter les réactions parasites lors de la synthèse automatique. Ce groupement peut être déprotégé dans les conditions de clivage standard. Le bras comporte un motif de type sérinol et l'alcool primaire est protégé par un groupement DMT. Ainsi, il va se comporter comme un phosphoramidite standard dans un cycle de déprotection/couplage en phase solide et permettre l'élongation de la séquence.

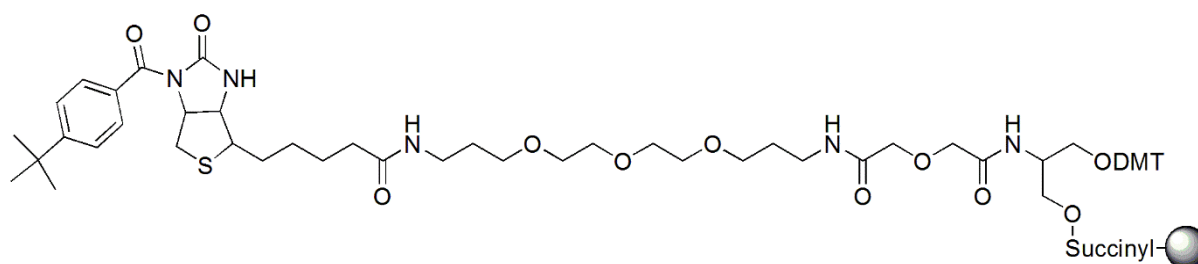


Figure 113. Structure de l'étiquette biotine liée à la colonne.

En ce qui concerne l'agent photoactivable, le nucléoside convertible que nous avons choisi d'utiliser est la 2-Fluoro-2'-désoxyinosine, qui est un analogue de la guanosine. Dans le contexte des duplex que nous souhaitons utiliser comme sonde, cela permet de placer la sonde photoactivable sur la base opposée à celle qui possède la base modifiée, soit à la position 16 du brin complémentaire. Le synthon phosphoramidite correspondant à ce nucléoside est disponible dans le commerce. Il est protégé sur sa position 6 par un groupement *para*-nitrophényléthyl. Ce dernier n'étant pas déprotégé dans les conditions de clivage de l'oligonucléotide de la colonne, l'utilisation de ce nucléoside convertible nécessite une étape de déprotection supplémentaire par une base, le 1,8-diazabicyclo[5.4.0]undéc-7-ène (DBU) qui intervient entre la fonctionnalisation post-synthétique et le clivage de la colonne. Par ailleurs, nous avons prévu d'introduire la ⁵I dU à la place de la thymidine à la position 17 du brin complémentaire en utilisant le phosphoramidite correspondant disponible également chez Glen Research (Figure 114).

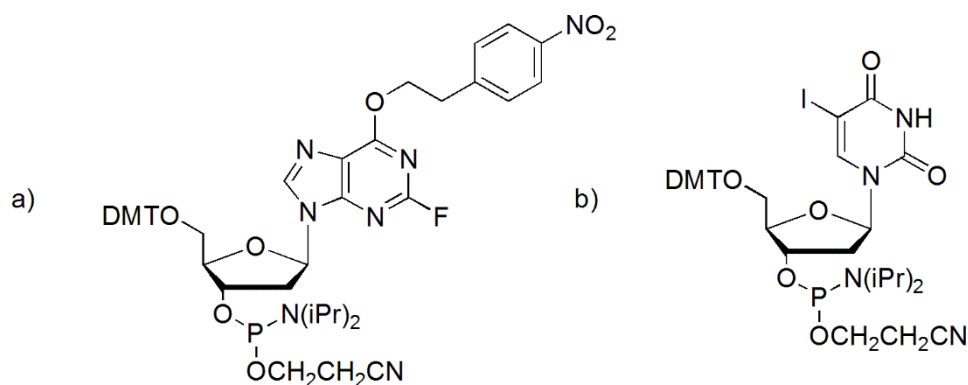


Figure 114. Phosphoramidite commerciaux utilisés ; a) 2-Fluoro-2'-désoxyinosine ; b) 5-Iodo-2'-désoxyuridine.

La méthode du « convertible nucleoside » implique, comme nous l'avons vu dans le chapitre précédent, d'utiliser une amine primaire, peu encombrée stériquement, pour permettre de fonctionnaliser une séquence d'ADN de manière post-synthétique.

2.2.1 Synthèse des amines comportant les groupements photoactivables

Nous avons choisi d'utiliser à cette fin un bras espaceur de type diamine comportant une chaîne alkyle flexible et modulable en taille. La modulation de la longueur de bras peut se faire facilement, ce qui permet de synthétiser des sondes photoactivables de longueurs différentes, qui peuvent permettre d'obtenir des informations différentes.

La synthèse des amines comportant les groupements photoactivables repose sur le couplage de type peptidique entre la diamine et une benzophénone ou une trifluorométhyldiazirine comportant une fonction acide carboxylique. La synthèse, effectuée en deux étapes, est présentée dans la figure 115. La première étape consiste en un couplage de la 1,6-hexyldiamine monoprotégée par un groupement Boc **56** avec la benzophénone **55** ou trifluorométhylphényldiazirine **59** préalablement activés en présence d'1-[Bis(diméthylamino)méthylène]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid-hexafluorophosphate (HATU) et de diéthylamine. Ces composés **57** et **60** sont ensuite déprotégés en présence d'acide trifluoroacétique dans le dichlorométhane pour obtenir, après

neutralisation, les amines primaires finales **58** et **61**, qui seront utilisées dans la fonctionnalisation post-synthétique.

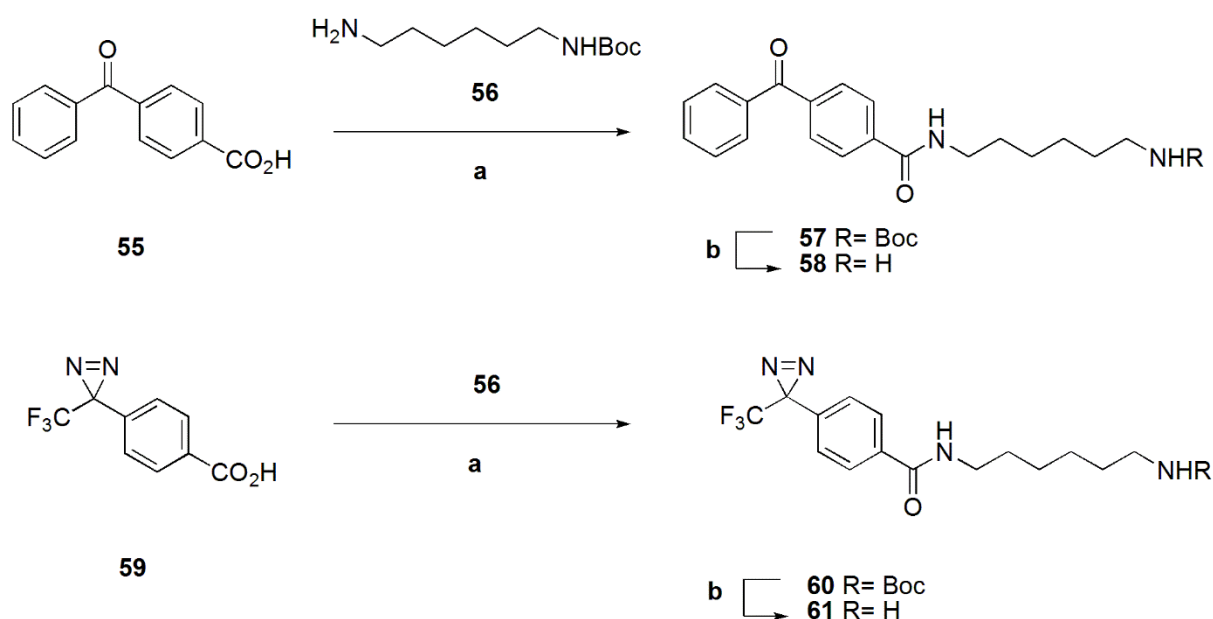


Figure 115. Synthèse des amines comportant un groupement photoactivable ; conditions et réactifs ; (a) *i*) HATU, DIEA, DMF *ii*) *tert*-butyl-*N*-(6-aminohexyl)carbamate 2, DIEA, DMF ; (b) TFA, CH_2Cl_2 .

2.2.2 Synthèse des oligonucléotides en phase solide

Nous avons mis en œuvre la synthèse automatique des oligonucléotides 30-mères contenant la 2-Fluoro-2'-désoxyinosine ou la 5-Iodo-2'-désoxyuridine sur un synthétiseur Expedite 8900 sur des colonnes de 0.2 μmol selon la stratégie DMT-on décrite précédemment. Nous avons réalisé ces synthèses à l'Institut de Chimie de Nice à l'Université de Nice-Sophia Antipolis dans l'équipe « Molécules Bioactives ».

Comme le montre la Figure 116, le rendement de couplage sur phase solide des nucléotides commerciaux est compris entre 95 et 100 % et le rendement obtenu lors du couplage du phosphoramidite de la 2F-dI ainsi que la ^5dU est comparable à ceux obtenus avec les phosphoramidites standards.

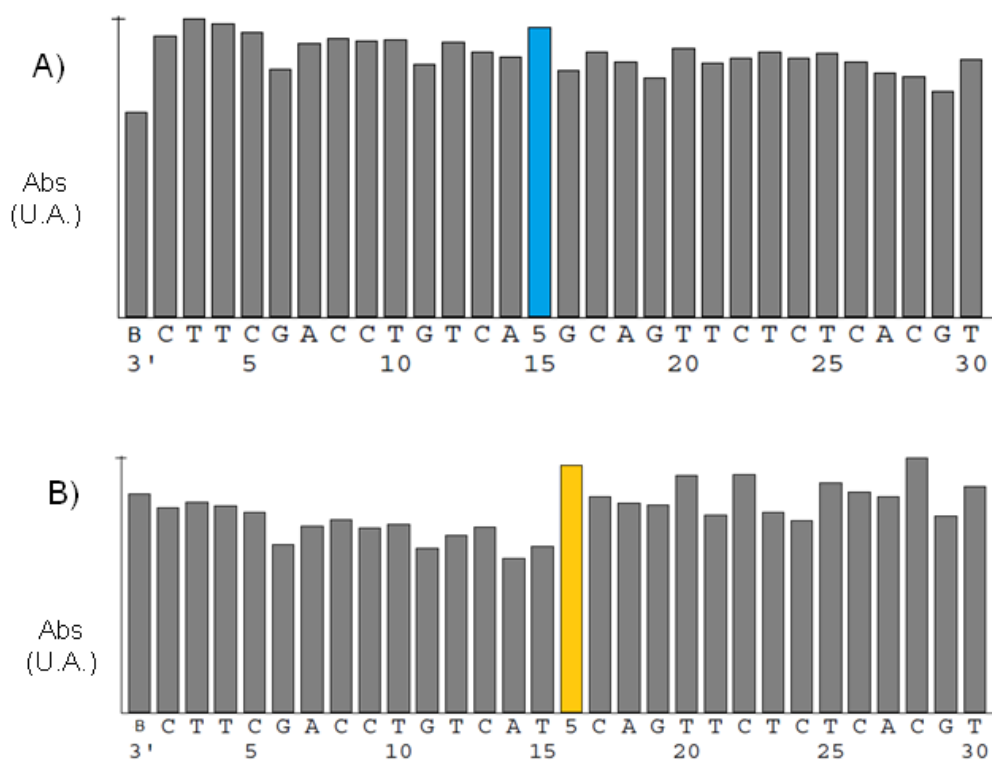


Figure 116. Rendements de couplage lors de la synthèse automatique (estimés à partir des profils d'absorption UV de déprotection des groupements DMT terminaux). ; A) La barre bleue correspond au couplage de la 5-Iodo-2'-désoxyuridine ; B) La barre orange correspond au couplage de la 2-Fluoro-2'-désoxyinosine. Dans les deux cas, la lettre B correspond à la déprotection de la biotine préchargée sur la colonne, la lettre 5 au couplage du phosphoramidite modifié.

En ce qui concerne les oligonucléotides comportant la ^{51}dU , ils ont été traités selon le même protocole que celui vu précédemment (déprotection à l'ammoniaque, purification par HPLC, détritulation suivie d'une deuxième purification par HPLC) et caractérisés par spectrométrie de masse MALDI-TOF.

Pour les oligonucléotides comportant la ^{2F}dI , une fois la synthèse terminée, les colonnes ont été séchées à l'argon puis conservées à -20°C en attendant la mise en œuvre du protocole de fonctionnalisation post-synthétique.

2.2.3. Fonctionnalisation post-synthétique des oligonucléotides

La mise au point des conditions de fonctionnalisation a été réalisée avec des séquences contrôles ne comportant pas de biotine en 3' et l'amine fonctionnalisée avec une

benzophénone. Le protocole de fonctionnalisation post-synthétique est le suivant : la colonne sur laquelle est greffée la séquence comportant une 2-Fluoro-2'-désoxyinosine est remise à température ambiante puis traitée avec une solution d'amine dans le DMSO. Le traitement se déroule à l'aide de deux seringues reliées à la colonne, dont les allers-retours réguliers permettent de faire passer la solution d'amine à travers la colonne. La colonne est conservée dans une étuve à la température indiquée pendant 24 à 48 h (Tableau 5).

Après le traitement avec la solution comportant l'amine, la colonne est ensuite lavée avec du DMSO puis de l'acétonitrile, traitée avec une solution de DBU dans l'acétonitrile de la même façon que précédemment et enfin lavée une dernière fois avec du méthanol et de l'acétonitrile. On procède ensuite au clivage de la séquence oligonucléotidique de la colonne avec une solution d'ammoniaque concentrée pendant la nuit, à température ambiante. L'oligonucléotide sous forme de dérivé 5'-*O*-diméthoxytrityle est ensuite traité comme précédemment décrit (purification par HPLC, détritylation et contrôle de la pureté par HPLC). Afin de vérifier que l'oligonucléotide obtenu comporte bien la modification souhaitée, sa masse est vérifiée par spectrométrie de masse MALDI-TOF. Ces analyses ont été réalisées par Vincent Guérineau à l'Institut de Chimie des Substances Naturelles, à Gif-sur-Yvette.

Le tableau 5 résume les conditions des différents essais ainsi que les résultats obtenus. Les deux premiers essais de fonctionnalisation n'ont pas permis d'obtenir l'oligonucléotide fonctionnalisé, les analyses par MALDI-TOF montrant que seule, la séquence non-fonctionnalisée était isolée. Nous avons donc opté pour des conditions plus drastiques, c'est-à-dire en augmentant la température, le temps de réaction ainsi que la concentration de la solution d'amine, permettant ainsi d'isoler le produit de fonctionnalisation avec un rendement de fonctionnalisation estimé à 85% à partir du profil HPLC.

Essai	Température	Temps	[Amine]	[DIEA]	Rendement fonctionnalisation
1	Ta	25	0.1 M	-	Non isolé
2	Ta	25	0.1 M	0.1 M	Non isolé
3	55	48	0.2 M	0.2 M	85 %

Tableau 5. Différentes conditions de fonctionnalisation de l'oligonucléotide testées.

Une fois les conditions de fonctionnalisation mises au point, deux séquences comportant la biotine à leur extrémité 3' et une 2-Fluoro-2'-désoxyinosine en position 16 ont été fonctionnalisées avec les amines comportant le motif benzophénone et le motif trifluorométhylediazirine présentés dans la figure 115. La figure 117 présente, à titre d'exemple, le spectre de masse ainsi que le profil HPLC de l'oligonucléotide fonctionnalisé avec la benzophénone.

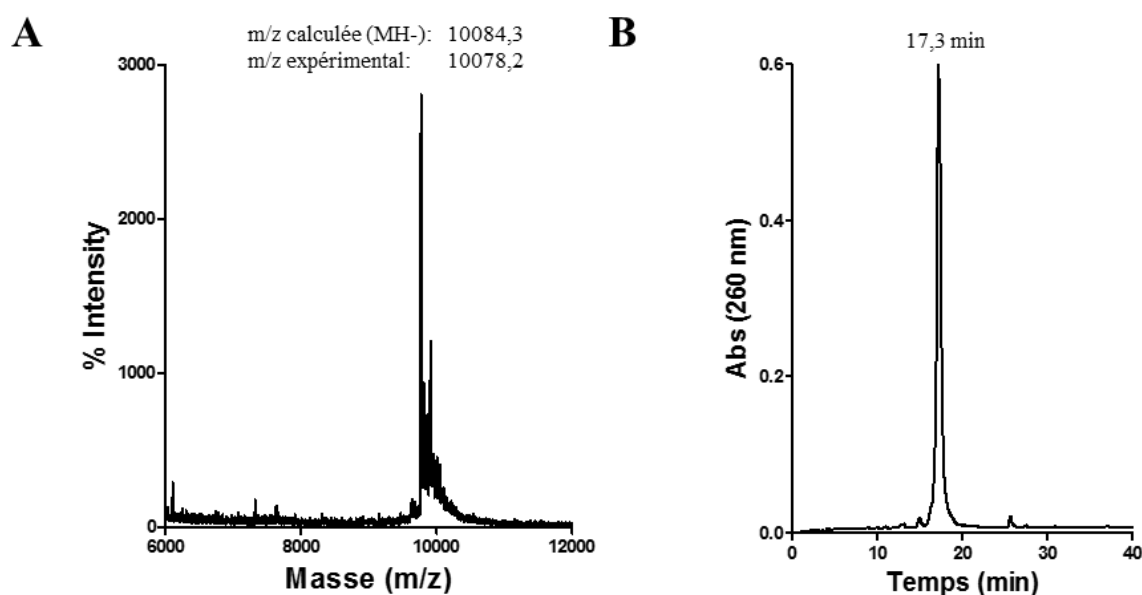


Figure 117. Caractérisation de l'oligonucléotide 16-Bzp-Biot ; A) Spectre de masse obtenu par MALDI-TOF ; B) Spectre HPLC de l'oligonucléotide DMT-off.

2.2.4. Récapitulatif des oligonucléotides synthétisés

En plus des oligonucléotides que nous avons synthétisés, nous avons également acheté des oligonucléotides comportant comme sonde photoactivable la ^{45}T à différentes positions chez Eurogentec. Les modifications chimiques introduites dans les différents oligonucléotides synthétisés ou achetés pour cette étude sont présentées dans la Figure 118.

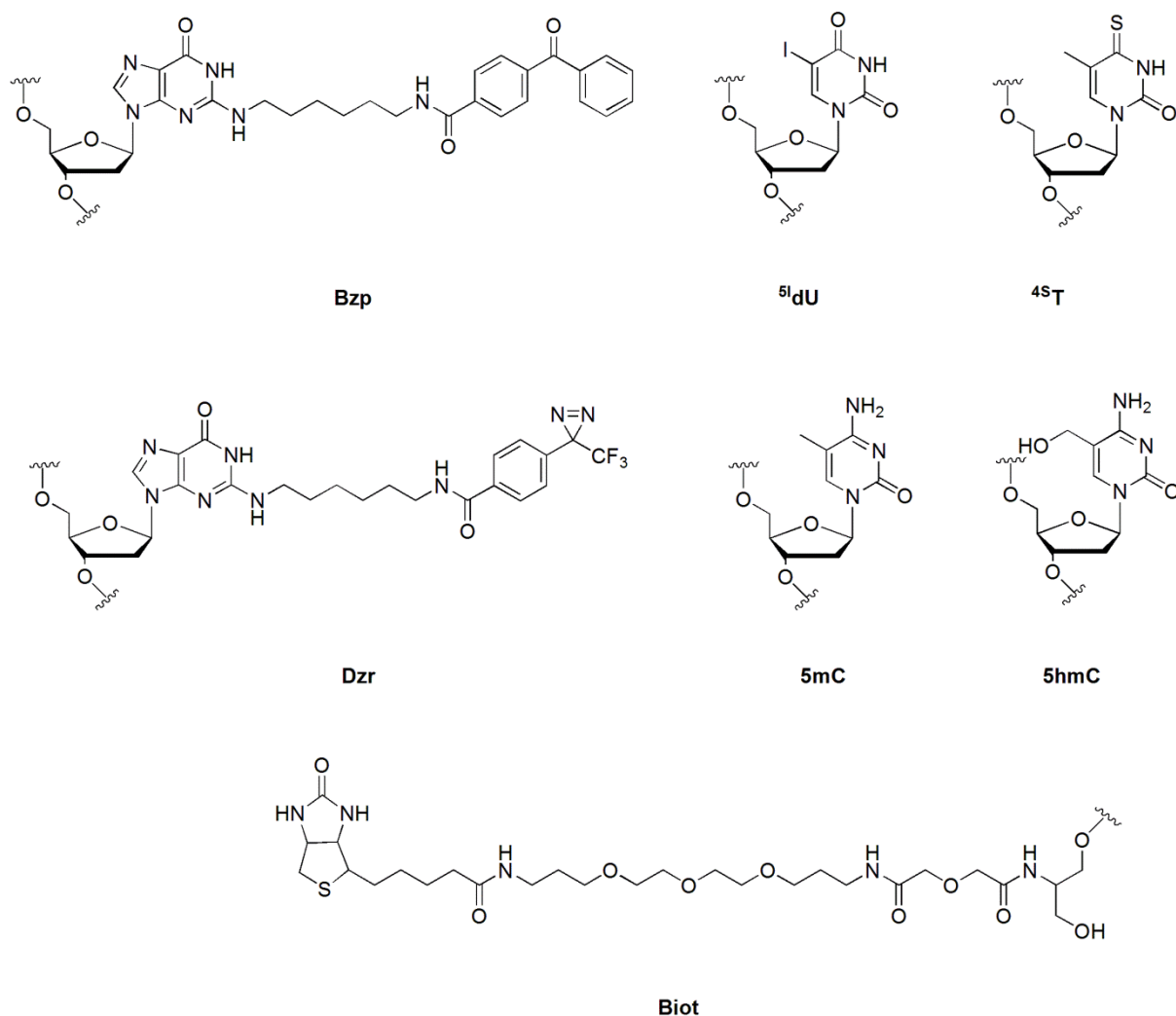


Figure 118. Structure des modifications introduites dans les oligonucléotides.

Afin d'identifier facilement chaque oligonucléotide modifié ou non, nous avons adopté la nomenclature suivante :

- dans le cas des brins portant les nucléosides photoactivables, chaque brin est identifié par la position où se trouve ce nucléoside en partant de l'extrémité 5', la nature de celui-ci, puis s'il porte une fonction biotine en 3' ou non (exemple : 16-Bzp-Biot). Le brin non modifié utilisé comme contrôle est nommé Ctrl-Biot ou Ctrl selon qu'il comporte une biotine ou non.

- Le brin portant (ou non) la cytosine modifiée est nommé C, M ou H selon qu'il porte respectivement en position 15 à partir de l'extrémité 5' une cytosine, une 5-méthylcytosine ou une 5-hydroxyméthylcytosine.

- On nomme les duplex en utilisant les noms des deux brins reliés par les lettres « ds » pour « double strand ». Par exemple, le duplex formé par l'hybridation des brins M et 16-Bzp-Biot sera nommé Mds-16-Bzp-Biot.

Ainsi, la liste des oligonucléotides obtenus ainsi que leur nom est présenté dans la Figure 119.

5' GAA GCT GGA CAG TAX GTC AAG AGA GTG CAA 3' 15	
X= 5hmC	H
X= 5mC	M
X= C	C

5' TTG CAC TCT CTT GAC XTA CTG TCC AGC TTC Biot 3' 16	
X= Bzp	16-Bzp-Biot
X= Dzir	16-Dzir-Biot
X= G	Ctrl-Biot

5' TTG CAC TCT CTT GAC GXA CTG TCC AGC TTC Biot 3' 17	
X= ^{4S}T	17-Iodo-Biot
X= ^{5I}dU	17-Thio-Biot

5' XTG CAC TCT CTT GAC GTA CTG TCC AGC TTC Biot 3' 1	
X= ^{4S}T	1-Thio-Biot

5' TTG CAC TCT CTT GAC XTA CTG TCC AGC TTC 3' 16	
X= G	Ctrl

5' TTG CAC TCT CTT GAC GXA CTG TCC AGC TTC 3' 17	
X= ^{5I}dU	17-Iodo
X= ^{4S}T	17-Thio

5' XTG CAC TCT CTT GAC GTA CTG TCC AGC TTC 3' 1	
X= ^{4S}T	1-Thio

Figure 119. Liste des sondes utilisées lors de ce chapitre et nomenclature associée.

Munis de ces oligonucléotides synthétiques, nous avons réalisé un certain nombre d'expériences préliminaires de mise au point visant d'une part à évaluer la stabilité des duplex modifiés et d'autre part à optimiser les conditions de photomarquage à l'aide de protéines modèles.

2.3 Etude de stabilité des duplexes étudiés

Nos oligonucléotides comportent plusieurs modifications chimiques sur les deux brins et nous avons vérifié dans un premier temps que leur présence ne déstabilisait pas la formation des duplexes. Ainsi, nous avons évalué leurs stabilités respectives par deux types d'expériences :

- d'une part, par des expériences de gel retard
- d'autre part, par des mesures de température de fusion de l'ADN.

2.3.1 Expériences de gel retard

Le principe de ces expériences est le suivant : le brin portant la sonde photoactivable est radiomarké en position 5' en présence de polynucléotide kinase et de [$\gamma^{32}\text{P}$]-ATP et purifié sur un gel d'acrylamide dénaturant (15% acrylamide, 7M urée). On ajoute à cet oligonucléotide dont la concentration est constante (10 nM) des concentrations croissantes de l'oligonucléotide complémentaire « froid ». Après avoir été chauffés à 95°C pendant 5 minutes, les échantillons sont laissés refroidir à température ambiante pendant au moins trois heures. Les échantillons sont ensuite séparés par migration électrophorétique sur un gel natif d'acrylamide 10% à 25°C. Le duplex formé étant plus lourd que le simple brin radiomarké, il va migrer plus lentement sur le gel d'acrylamide. La figure 120 représente les autoradiogrammes de différentes expériences avec les duplex comportant les quatre agents de photomarquage ainsi qu'un duplex contrôle non modifié.

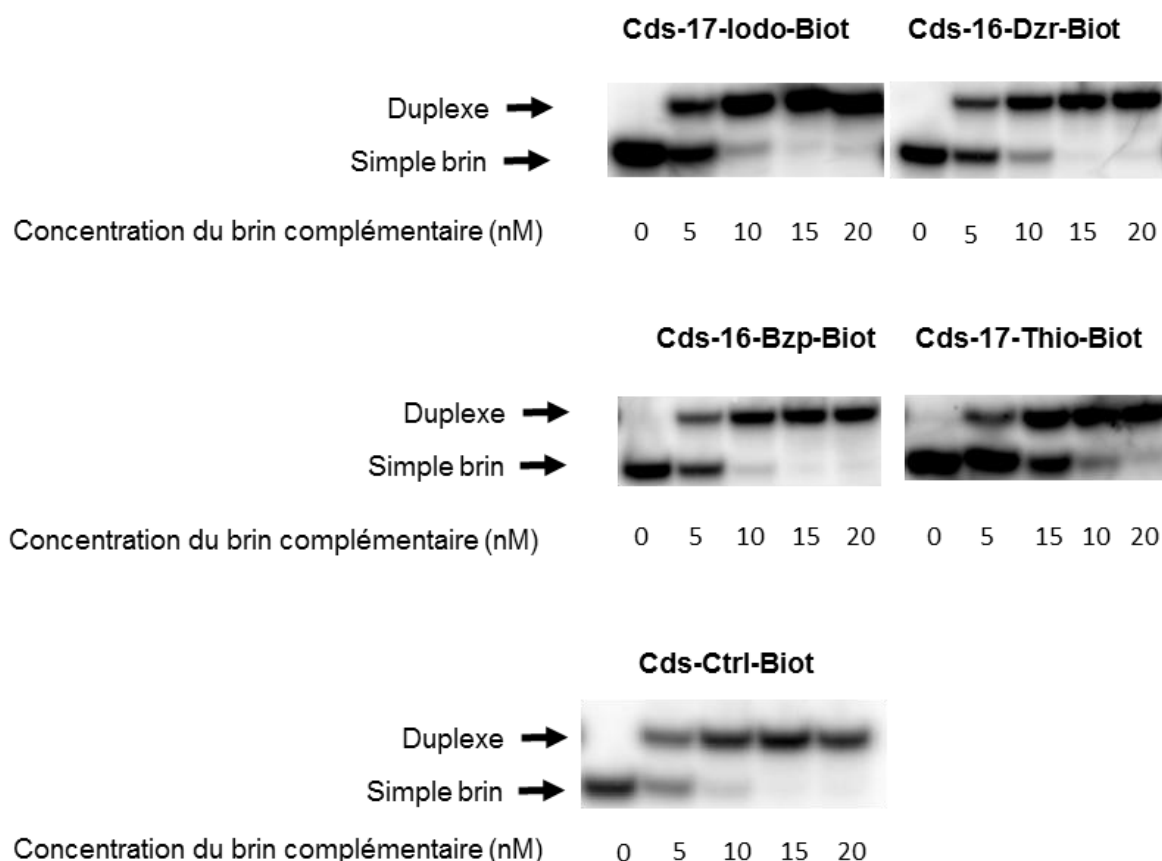


Figure 120. Etude de la stabilité des différents duplex par retard sur gel.

A partir de la mesure des quantités de complexe retardé par rapport à celles des bandes non retardées aux différentes concentrations, des valeurs de C_{50} ont été calculées pour chaque duplex et sont reportées dans le tableau 6.

Duplex	C_{50} (nM)
Cds-16-Bzp-Biot	6.5 ± 1.1
Cds-16-Dzr-Biot	5.8 ± 0.2
Cds-17-Iodo-Biot	4.8 ± 0.9
Cds-17-Thio-Biot	9.0 ± 1.5
Cds-Ctrl-Biot	4.7 ± 0.5

Tableau 6. Valeurs des C_{50} obtenues pour les différentes sondes. C_{50} est la concentration d'ADN complémentaire à laquelle 50% du duplex est formé dans les expériences de gel retard.

Le duplex contrôle (Cds-ODN-Biot) présente comme attendu une C_{50} autour de 5nM, la concentration du brin radiomarké étant de 10 nM. La présence de l'atome d'iode à la position C5 de l'uridine ne semble pas avoir de conséquence sur la formation du duplex, ce qui est cohérent avec la similarité de l'encombrement stérique de l'iode et du groupement méthyle. L'introduction d'un groupement benzophénone ou diazirine déstabilise partiellement

la formation du duplex, l'effet étant plus marqué avec la benzophénone. Étonnamment, l'introduction d'une 4-Thio-2'-désoxythymidine semble déstabiliser de façon plus importante la formation de duplex. Cette déstabilisation a déjà été rapportée par d'autres groupes de recherche²²³.

2.3.2 Expériences de température de fusion

La dénaturation thermique d'un duplex d'ADN entraîne un changement des propriétés spectroscopiques de celui-ci, et se traduit notamment par une absorbance différente. Si l'on chauffe progressivement une solution d'ADN double brin et si l'on suit la densité optique à 260 nm en fonction de la température, on constate une augmentation importante de la densité optique comprise entre 20 et 30%. Cet effet est appelé effet hyperchrome. Il est dû à la séparation des deux brins de la double-hélice par rupture des empilements des bases et des liaisons hydrogène interchaînes : on parle de fusion de l'ADN.

Cette propriété a été exploitée pour mesurer un paramètre de l'ADN, sa température de fusion, ou T_m . La courbe de densité optique en fonction de la température a une forme sigmoïdale ; elle permet de déterminer un point de transition, la T_m , qui correspond à la température à laquelle les molécules d'ADN sont à demi-dénaturées (température de demi-dissociation, Figure 121). La T_m dépend de la longueur du polymère, de la composition en bases (plus la teneur en paires GC d'un ADN est grande, plus la T_m est élevée) et de la force ionique des solutions (la présence de cations stabilise l'ADN, ce qui induit une augmentation de la T_m).

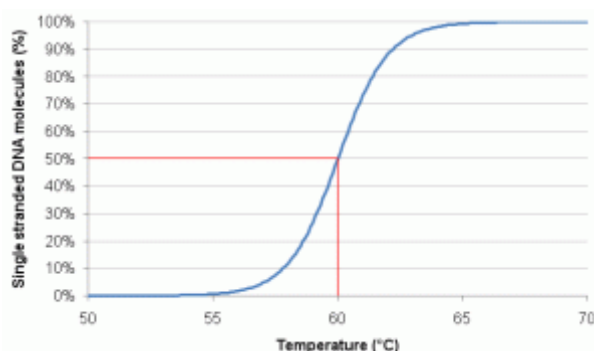


Figure 121. Courbe caractéristique permettant de définir la T_m d'un duplex d'ADN (Tiré de <http://www.entelechon.com/2008/08/dna-melting-temperature/>).

Le tableau 7 récapitule les températures de fusion des différents duplex que nous avons mesurées, afin d'évaluer l'effet de l'introduction de nos modifications sur la stabilité des sondes.

- L'introduction d'une biotine à l'extrémité 3' d'un des brins du duplex n'a pas d'effet stabilisant ou déstabilisant sur la formation du duplex, comme on peut le constater en comparant les T_m des duplex Cds-17-Iodo-Biot (73.9°C) et Cds-17-Iodo (73.7°C).

- La comparaison du duplex contrôle Cds-Ctrl avec les duplex comportant un groupement photoactivable indique que l'introduction d'une trifluorométhylphényldiazirine, d'une ^{45}T ou d'un ^{51}dU ne déstabilise pas les duplex formés. En effet, leurs T_m sont très proches, la différence maximale ne dépassant pas 0.3°C. Cependant, l'introduction d'une benzophénone a un léger effet déstabilisateur: la T_m de Cds-16-Bzp-Biot diminue de 1.7°C par rapport à celle du contrôle.

- L'introduction d'une 5mC ou d'une 5hmC dans un duplex standard ne déstabilise pas celui-ci, comme on peut le constater en comparant les T_m des duplex Cds-Ctrl, Mds-Ctrl et Hds-Ctrl. Ces résultats sont en accord avec les données de la littérature³⁷.

Duplex	T_m (°C)
Cds-16-Bzp-Biot	72
Cds-16-Dzr-Biot	73.8
Cds-17-Iodo-Biot	73.9
Cds-17-Thio-Biot	73.4
Cds-Ctrl	73.7
Hds-Ctrl	73.7
Mds- Ctrl	73.6
Cds-17-Iodo	73.7

Tableau 7. Températures de fusions de différents duplex utilisés dans la suite de l'étude

Les deux séries d'expériences pour évaluer la stabilité des duplex convergent donc à l'exception du cas de la ^{45}T qui, dans le cas de l'expérience de retard sur gel semblait indiquer une légère déstabilisation. Les modifications que nous avons introduites dans nos oligonucléotides impactent peu la stabilité des duplex et semblent appropriées pour des applications de photomarquage.

Nous avons ensuite cherché à mettre au point les conditions de photomarquage avec une protéine modèle.

2.4 Mise au point des conditions de photomarquage avec une protéine modèle

2.4.1 Etude des complexes duplex-DNMT3a

Notre choix s'est porté dans un premier temps sur une protéine déjà étudiée au laboratoire, le domaine catalytique de la DNA méthyltransférase 3a murine (Dnmt3a), composé des acides aminés 608 à 908. Ce choix s'est effectué pour les raisons suivantes : cette protéine est capable de méthylater la séquence que nous avons choisie au niveau du dinucléotide CpG, et par conséquent d'interagir avec notre séquence ; par ailleurs, nous avons la possibilité de la produire et de la purifier aisément sous forme de protéine recombinante en grande quantité au laboratoire. La Dnmt3a assure une méthylation *de novo*, c'est-à-dire qu'elle a une préférence pour des duplex non méthylés par rapport à des duplex hémiméthylés.

Nous avons tout d'abord cherché à trouver les meilleures conditions de complexation entre notre duplex et la Dnmt3a. L'étude a été réalisée avec le premier duplex photoactivable auquel nous avons eu accès, à savoir le duplex Cds-17-Thio. Ce duplex a été incubé à une concentration de 5 nM avec différentes quantités de Dnmt3a en présence ou non du cofacteur de l'enzyme, la S-adenosyl-L-méthionine (AdoMet) pendant une heure à 37°C. Les échantillons ont ensuite été séparés sur un gel natif d'acrylamide 10%. (Figure 122). Quand un complexe se forme entre notre duplex et Dnmt3a, il va migrer plus lentement sur le gel d'acrylamide, ce qui va permettre de quantifier le pourcentage de complexe duplex/Dnmt3a formé en fonction de la concentration de protéine.

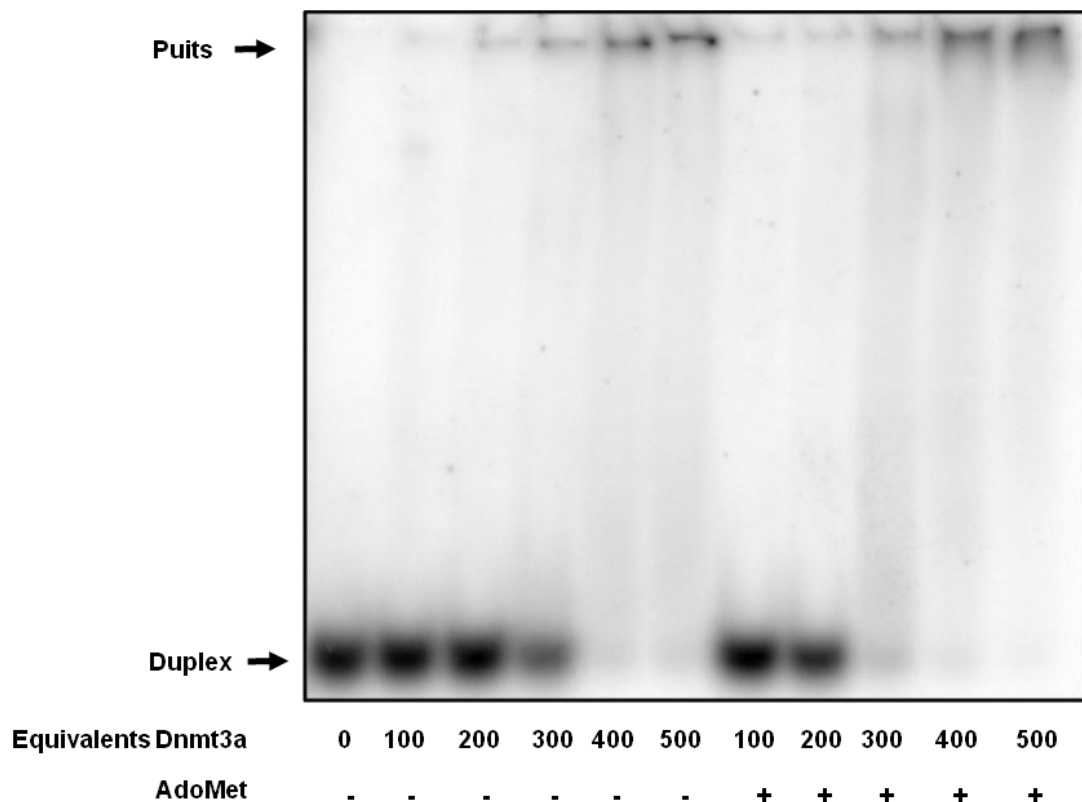


Figure 122. Complexation de la Dnmt3a avec Cds-17-Thio ; séparation sur gel natif d'acrylamide (Autoradiogramme).

On observe bien une interaction entre Dnmt3a et la sonde Cds-17-Thio. Ainsi, plus la concentration de la protéine augmente, plus l'intensité de la bande correspondant au duplex non complexé diminue. De plus, la protéine Dnmt3a possède une affinité plus importante pour le duplex lorsque le cofacteur AdoMet est présent. Cependant, l'affinité de l'enzyme pour le duplex est très faible : il faut utiliser entre 300 et 500 équivalents de Dnmt3a pour complexer entièrement le duplex. Enfin, la séparation du complexe Dnmt3a-duplex n'est pas satisfaisante. En effet, le complexe reste en majorité dans les puits du gel ou forme des trainées non résolues, ce qui pose des problèmes tant pour la quantification des complexes que pour l'évaluation de leur nature, la stoechiométrie étant difficilement évaluable dans ces conditions.

Différentes conditions de séparation sur gel d'acrylamide ont été testées sans réussir à faire migrer les complexes formés hors des puits. L'utilisation de gel d'agarose n'a pas apporté d'amélioration de la résolution des gels (Figure 123).

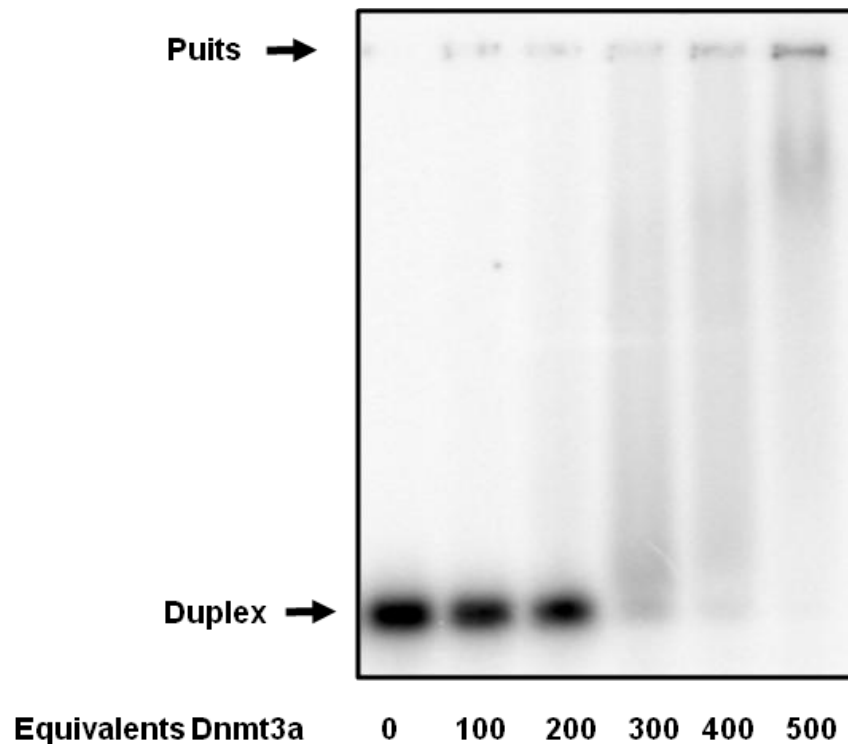


Figure 123. Complexation entre Dnmt3a et Cds-17-Thio ; séparation sur gel natif d'agarose (Autoradiogramme).

Malgré cela, nous avons réalisé des essais de photomarquage entre la Dnmt3a et le duplex Cds-17-Thio. L'incubation du duplex à une concentration de 5 nM avec différentes quantités de DNMT3a en présence de son cofacteur AdoMet pendant 1 heure à 37°C a ainsi été suivie de l'irradiation des échantillons avec une lampe UV à 365 nm pendant 45 minutes (12W, Bioblock Scientific). L'analyse des échantillons sur gel dénaturant a permis d'observer la formation d'un complexe entre la sonde oligonucléotidique et l'enzyme (Figure 124), avec un maximum de 12% de rendement de photomarquage pour l'échantillon incubé avec 500 équivalents de DNMT3a.

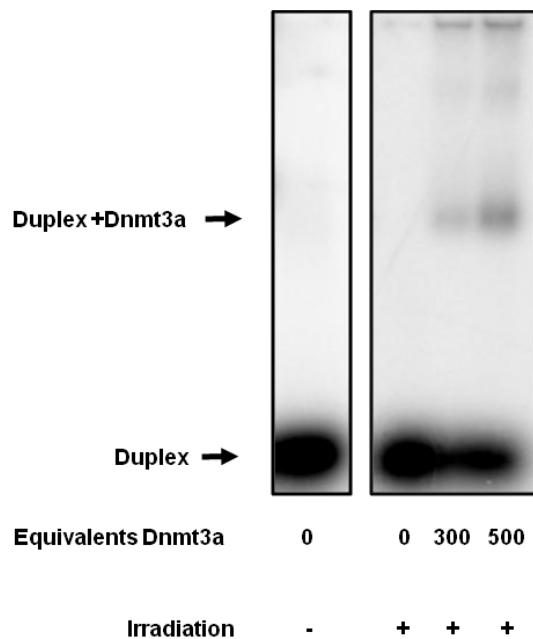


Figure 124. Photomarquage de DNMT3a par la sonde Cds-17-Thio. Séparation sur gel natif d'acrylamide (Autoradiogramme).

En conclusion, la Dnmt3a complexe nos sondes et il est possible de la photomarquer. Néanmoins, la faible affinité de l'enzyme pour les duplex combinée à l'analyse difficile des complexes formés par gel-retard nous ont conduits à mettre au point nos conditions de photomarquage à l'aide d'une autre protéine modèle.

2.4.2 Etude des complexes simple brin-RPA

Pour cela, nous avons fait le choix de travailler sur la protéine de réplication A humaine (hRPA), une protéine largement étudiée dans l'équipe de Carole Saintomé au Muséum National d'Histoire Naturelle avec qui nous avons collaboré dans le cadre de cette étude. La hRPA est la principale protéine de liaison au simple brin humaine. C'est un hétérotrimère constitué de sous-unités de 70, 32 et 14 kDa, nommées respectivement hRPA1, hRPA2 et hRPA3. Elle a été découverte en 1988, lors de l'identification des facteurs de transcription nécessaires à la réplication du virus simien SV40, un modèle de réplication eucaryote²⁵⁸. Elle intervient dans un grand nombre de processus du métabolisme de l'ADN tels que la réparation, les points de contrôle de l'ADN endommagé, la régulation de la transcription et le maintien des télomères. Dans la majorité des cas, hRPA s'associe aux

régions simple brin créés durant ces différents processus et les stabilise, mais interagit aussi avec les protéines du métabolisme de l'ADN afin de réguler leur activité.

De même que la Dnmt3a, nous pouvions produire et purifier cette protéine en grande quantité. Par ailleurs, sa forte affinité pour les simples brins sans spécificité de séquence nous a semblé intéressante afin d'effectuer une étude comparative des différents agents de photomarquage incorporés dans nos oligonucléotides. En effet, nous avons vu que chacun des groupements possède une photochimie et une réactivité qui lui est propre. L'évaluation des propriétés des sondes contenant une benzophénone, une diazirine, la ^5I dU ou la ^{48}T permettra de mettre à jour leurs avantages et inconvénients dans l'étude des interactions entre ADN et protéines.

Tout d'abord, nous avons cherché à évaluer les affinités respectives de nos oligonucléotides comportant les agents de photomarquage pour la protéine hRPA. Les sondes ont été incubées pendant 15 minutes avec des concentrations croissantes de protéine, puis ont été analysées sur un gel natif d'acrylamide (5% AA 19 :1, 0.5X TBE). Les différents gels sont présentés dans la figure 125.

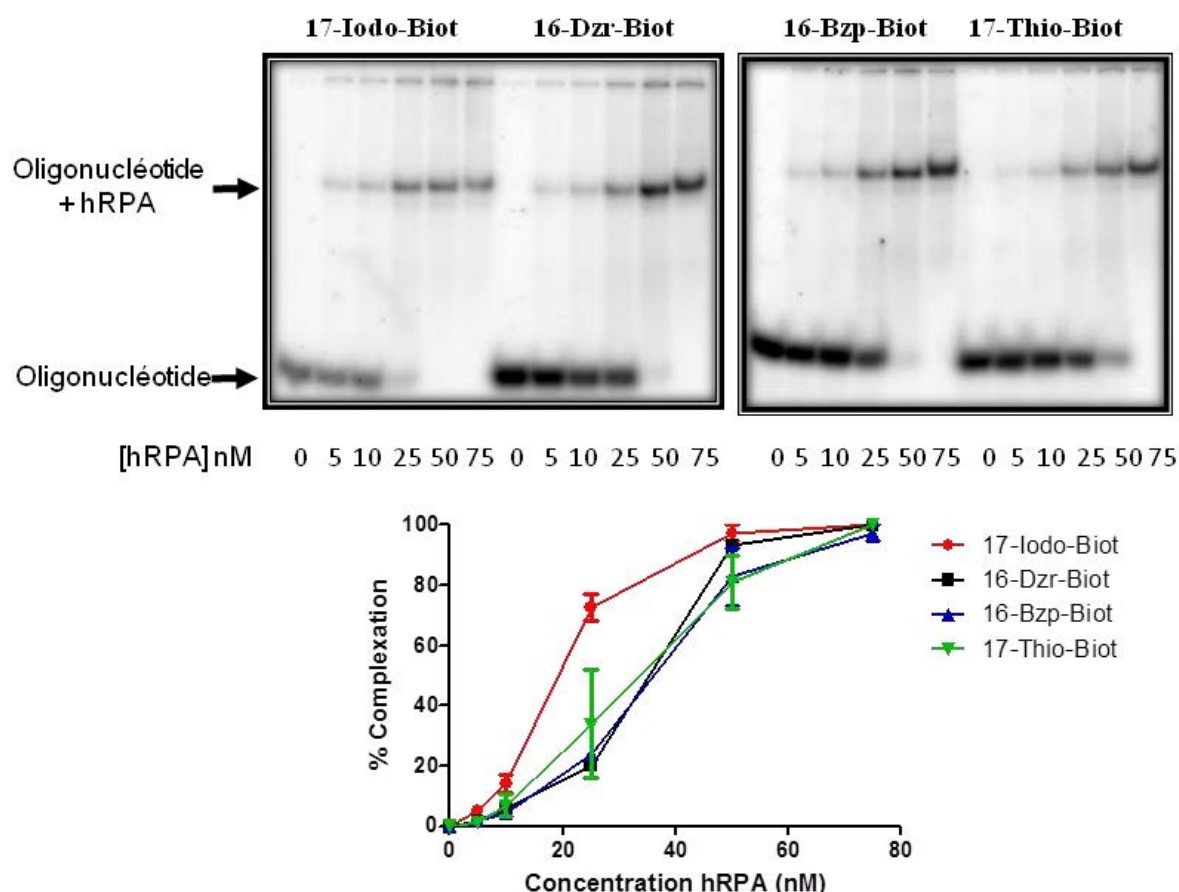


Figure 125. Interaction entre hRPA et oligonucléotide simple brin (Autoradiogramme). Le pourcentage de complexation a été calculé en comparant l'intensité de la bande correspondant oligonucléotide-hRPA par rapport à l'intensité totale (oligonucléotide complexé et libre).

On peut constater que hRPA a la plus grande affinité pour la sonde contenant la Cds-17-Iodo-Biot, tandis qu'elle possède une affinité similaire pour les trois autres sondes. De plus, cette expérience nous a permis de trouver des conditions de liaison où toutes les sondes oligonucléotidiques sont complexées par hRPA. Ce sont ces conditions (7.5 eq, 75 nM hRPA) que nous allons utiliser pour évaluer les propriétés de photomarquage des différentes sondes.

Comme nous l'avons mentionné, nous avons d'abord choisi RPA comme protéine modèle pour sa forte affinité pour le simple brin qui nous permettrait de faire une mise au point des conditions de photomarquage avec nos différents oligonucléotides. Ayant, cette protéine en main, nous nous sommes également intéressés à étudier son interaction potentiellement privilégiée avec l'autre brin, celui comportant la 5hmC. En effet, une publication précédente a montré que la protéine RPA du Xénope reconnaissait

préférentiellement des séquences contenant des bases oxydées de la thymidine, comme la 5hmU ou la 5fU, par rapport à des séquences contenant les mêmes bases non modifiées²⁵⁹. Les séquences d'ADN C et H (contenant respectivement une cytosine ou une 5hmC en position 15) ont donc été incubées avec des concentrations croissantes de hRPA dans les conditions préalablement décrites. Après séparation sur un gel d'acrylamide 5%, la complexation entre hRPA et l'ADN a été évaluée (Figure 126).

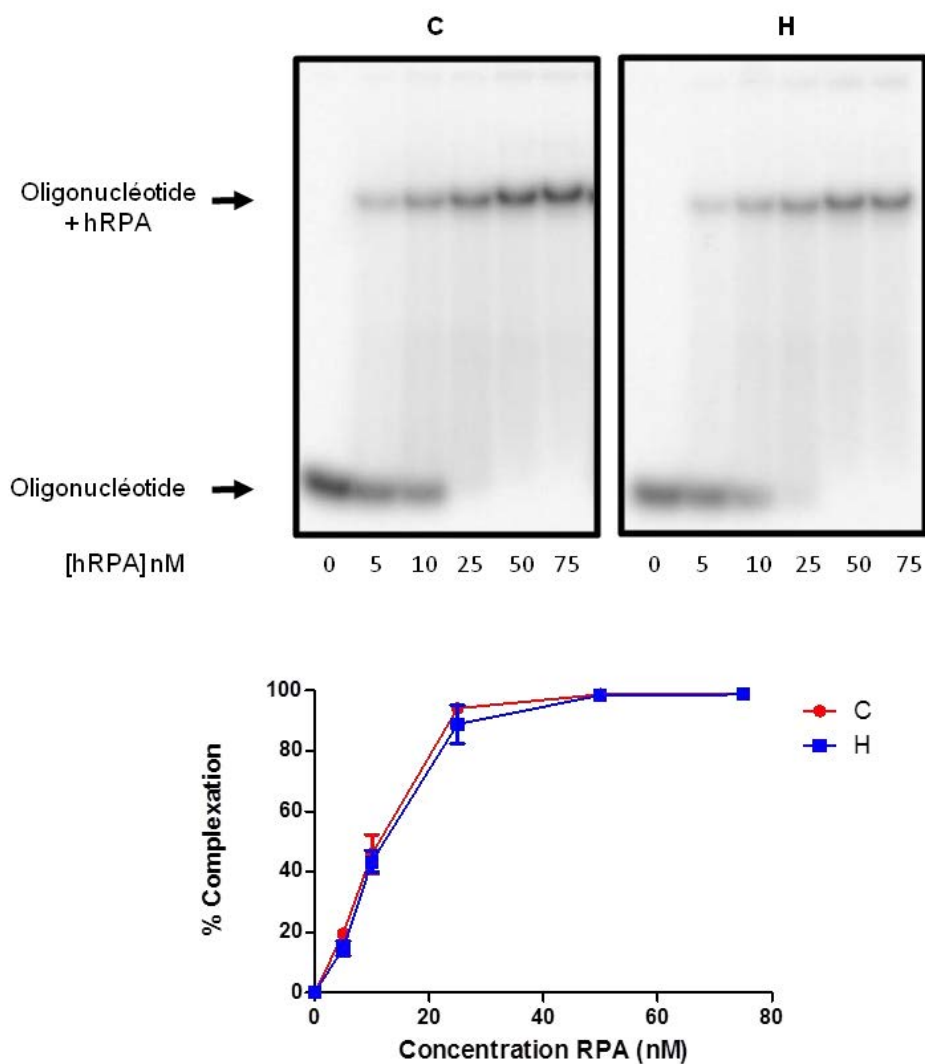


Figure 126. Complexation des brins H ou C en présence de la protéine hRPA (Autoradiogramme). Le pourcentage de complexation a été calculé en comparant l'intensité de la bande correspondant au complexe oligonucléotide-hRPA par rapport à l'intensité totale (oligonucléotide complexé et libre).

Pour notre séquence simple brin modèle, nous n'avons pas observé de différence de complexation par hRPA entre le brin contenant une cytosine et celui contenant une 5hmC. La différence avec l'étude précédente pourrait provenir soit de la nature de la protéine RPA (humaine/xénope), soit de la séquence, soit de la base (5hmU/5hmC). Les auteurs de l'étude mentionnée expliquaient l'affinité préférentielle de RPA pour les dérivés oxydés de la Thymidine comme une reconnaissance des dégâts oxydatifs de l'ADN. Nous nous sommes donc demandés si, dans le contexte d'un mécanisme de réparation, hRPA ne pouvait pas reconnaître la 5hmC dans de l'ADN double brin. Cependant, des expériences préliminaires de retard sur gel avec les duplex Cds-Ctrl, Mds-Ctrl et Hds-Ctrl n'ont pas permis de faire apparaître de différence dans la liaison de hRPA avec le duplex contenant la 5hmC et ceux contenant la 5mC et la C.

Les différentes sondes oligonucléotidiques obtenues ont des propriétés photophysiques et photochimiques différentes, comme présenté dans l'introduction de ce chapitre. Afin d'évaluer leur comportement dans des réactions de photomarquage et de choisir la sonde la plus efficace, nous avons réalisé une étude cinétique. Chacune des sondes a été placée dans la condition de complexation où 100% de la sonde oligonucléotidique était en complexe avec hRPA, déterminées dans l'expérience précédente. Après incubation à température ambiante, les échantillons ont été irradiés pendant deux heures avec la lampe UV précédemment présentée. Toutes les 20 minutes, un prélèvement est effectué et mis à l'abri de la lumière. Après migration de ces échantillons sur un gel d'acrylamide dénaturant, on peut évaluer les rendements de photomarquage (Figure 127).

Afin de vérifier que la formation de complexes covalents observés était bien le résultat de l'activation des groupements photoactivables et non d'une réactivité non spécifique des acides nucléiques, un oligonucléotide sans agent de photomarquage a été irradié dans les mêmes conditions (Figure 127). On constate qu'au bout de deux heures d'irradiation, le brin d'ADN ne comportant pas de groupement photoactivable n'a pas formé de complexe retardé avec hRPA. La formation des adduits covalents soligonucléotide-hRPA est bien due à la réactivité des groupements photoactivables et non à une réactivité non spécifique des acides nucléiques naturels.

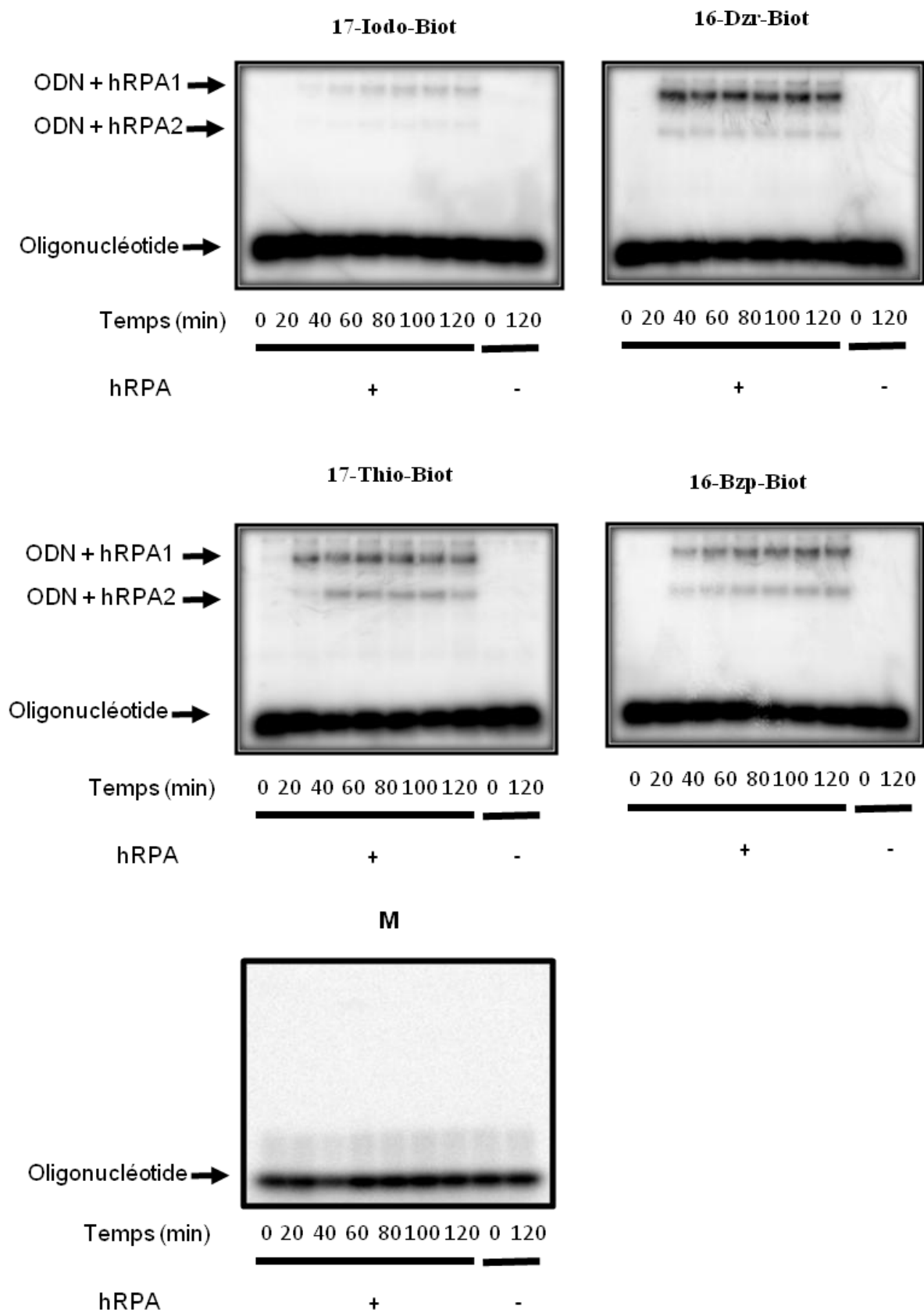


Figure 127. Cinétique d'irradiation des différentes sondes en présence de hRPA (Autoradiogramme, séparation par gel d'acrylamide dénaturant).

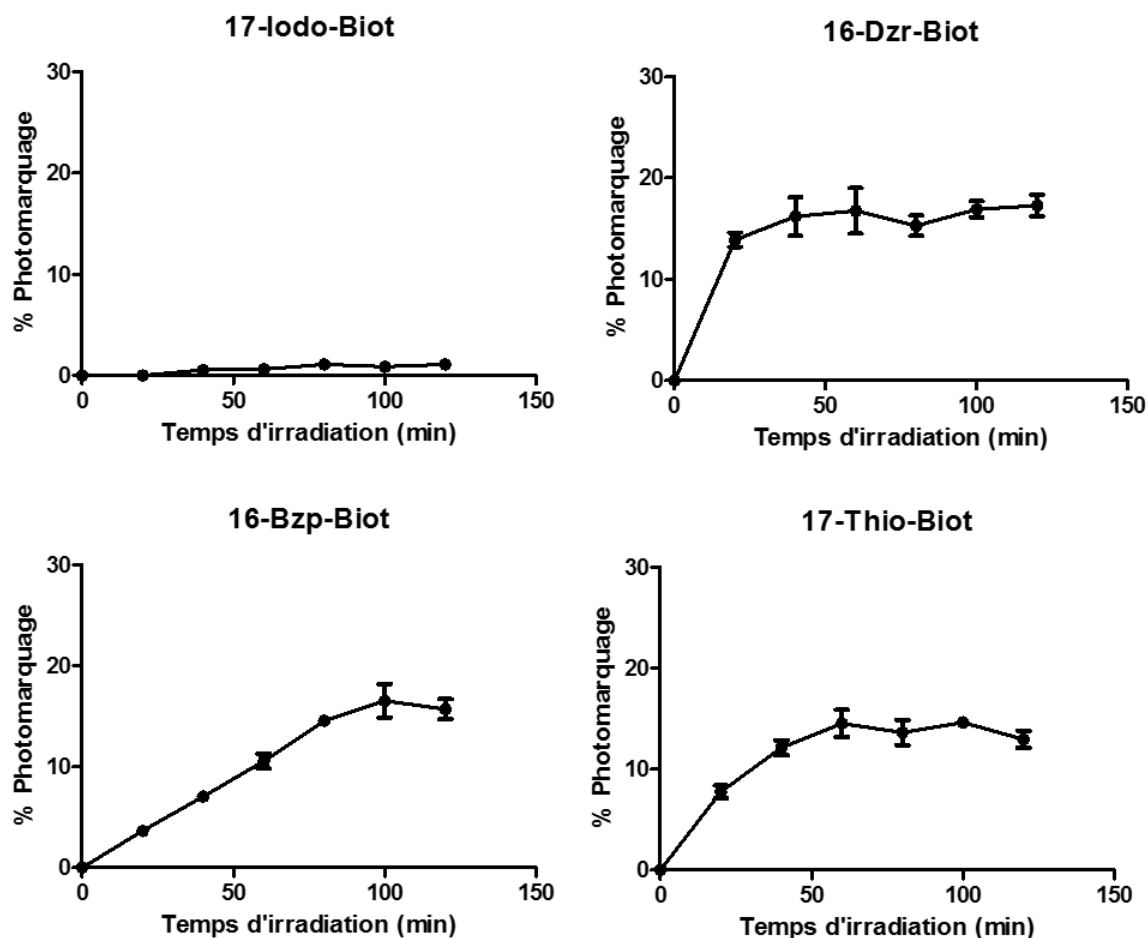


Figure 128. Evolution du pourcentage de photomarquage au cours du temps en fonction de la nature de la sonde. Exploitation des résultats présentés dans la figure 127. Le pourcentage de photomarquage a été calculé en comparant l'intensité des bandes correspondant aux complexes covalents oligonucléotide-hRPA1 et 2 par rapport à l'intensité totale (oligonucléotide lié covalamment et libre).

L'évolution des rendements (ou pourcentages) de photomarquage est présentée dans la figure 128. La ^5I dU est très peu efficace, avec un rendement de photomarquage maximal au bout de 2 heures de 3%. La lampe UV utilisée n'est donc pas optimale pour irradier cette base.

Les sondes contenant une diazirine, une benzophénone et la ^{45}T donnent toutes lieu à des rendements de photomarquage important. La sonde 16-Dzr-Biot atteint son rendement maximal de photomarquage au bout de 20 minutes, ce qui est cohérent avec la formation d'un carbène très réactif, qui va soit former un adduit covalent avec la protéine à proximité soit être inactivé par des réactions parasites, notamment avec l'eau. Le rendement de photomarquage de la sonde 16-Bzp-Biot augmente progressivement au cours du temps, pour atteindre un

maximum au bout de 100 minutes. Cela s'explique par l'équilibre entre le diradical et la cétone, qui permet d'améliorer le rendement de photomarquage avec une irradiation plus longue. La sonde 17-Thio-Biot présente elle aussi une dépendance de son rendement de photomarquage au cours du temps mais moins flagrante qu'avec la benzophénone. Cela peut s'expliquer par le fait que la plage de longueur d'onde de la lampe UV utilisée pour cette étude n'est pas optimale pour l'excitation de la 4-thioT, dont le maximum d'absorption se situe à 330 nm. Le rendement maximal de photomarquage est atteint au bout d'une heure, soit plus rapidement que pour la benzophénone.

Parmi les quatre sondes photoactivables, la diazirine est la plus efficace, avec un rendement maximum de 19% atteint dès 20 minutes d'irradiation. La benzophénone atteint des rendements de photomarquage comparables mais au bout d'un temps d'irradiation beaucoup plus long, ce qui présente des inconvénients (dégradation de l'échantillon irradié à cause de la chaleur dégagée par exemple). La ^{45}T est moins efficace, le rendement de photomarquage maximal atteint étant de 13%. La ^{51}dU ne donne quasiment pas de photomarquage. Pour remédier à l'inadéquation de la lampe UV utilisée pour cette sonde photoactivable, et tenter d'optimiser les rendements des autres sondes, des tentatives d'irradiation utilisant des diodes électroluminescentes à des longueurs d'ondes spécifiques ont été réalisées et seront présentées plus tard dans ce chapitre.

On peut remarquer aussi des différences dans le photomarquage des sous-unités de hRPA. Si hRPA1 est à chaque fois l'unité préférentiellement photomarkuée, la sonde contenant une benzophénone marque plus la seconde sous-unité, hRPA2, comparativement à la ^{51}dU et la diazirine. Ces différences dans le photomarquage des sous-unités sont sûrement dues à des différences de positionnement des groupements photoactivables dans l'interaction avec hRPA. Ceci peut permettre d'obtenir des informations intéressantes lors de l'étude de l'interaction entre nos sondes et les protéines et ces différentes propriétés mériteront d'être exploitées ultérieurement.

2.4.3 Essais d'irradiation avec des LEDs

Afin de remédier au très faible rendement de photomarquage de l'oligonucléotide comportant la ^5I dU, des essais de photomarquage avec des diodes électroluminescentes (LEDs) ont été réalisées. Les LEDs sont des sources de rayonnement moins puissantes que la lampe UV utilisée dans les études de photomarquage, mais émettant sur une plage de longueur d'onde beaucoup plus resserrée. La faible chaleur dégagée par ce type d'irradiation constitue aussi un avantage de ce type de dispositif. Des expériences de photomarquage d'affinité utilisant la ^{48}T et la ^5I dU se basant sur des LEDs ont déjà été publiés dans la littérature²⁶¹. Il nous a donc paru intéressant d'utiliser ce dispositif pour essayer d'améliorer le rendement de photomarquage de la ^5I dU, ainsi que d'étudier l'irradiation de nos autres agents photoactivables en présence de ces LEDs, afin de voir si nous pourrions gagner en efficacité de photomarquage du fait de la spécificité de la longueur d'onde utilisée.

Nous nous sommes donc équipés chez le fournisseur Thorlabs de LEDs émettant à deux longueurs d'onde différentes: 300 et 365 nm. La LED à 300 nm permet d'irradier la ^5I dU, tandis que la LED à 365 nm permet d'irradier la benzophénone, la diazirine et dans une moindre mesure la ^{48}T . D'après le constructeur, elles possèdent une largeur maximale à mi-hauteur du pic d'émission de 15nm. La figure 129 présente les spectres d'émission standard des deux diodes différentes fournis par le constructeur.

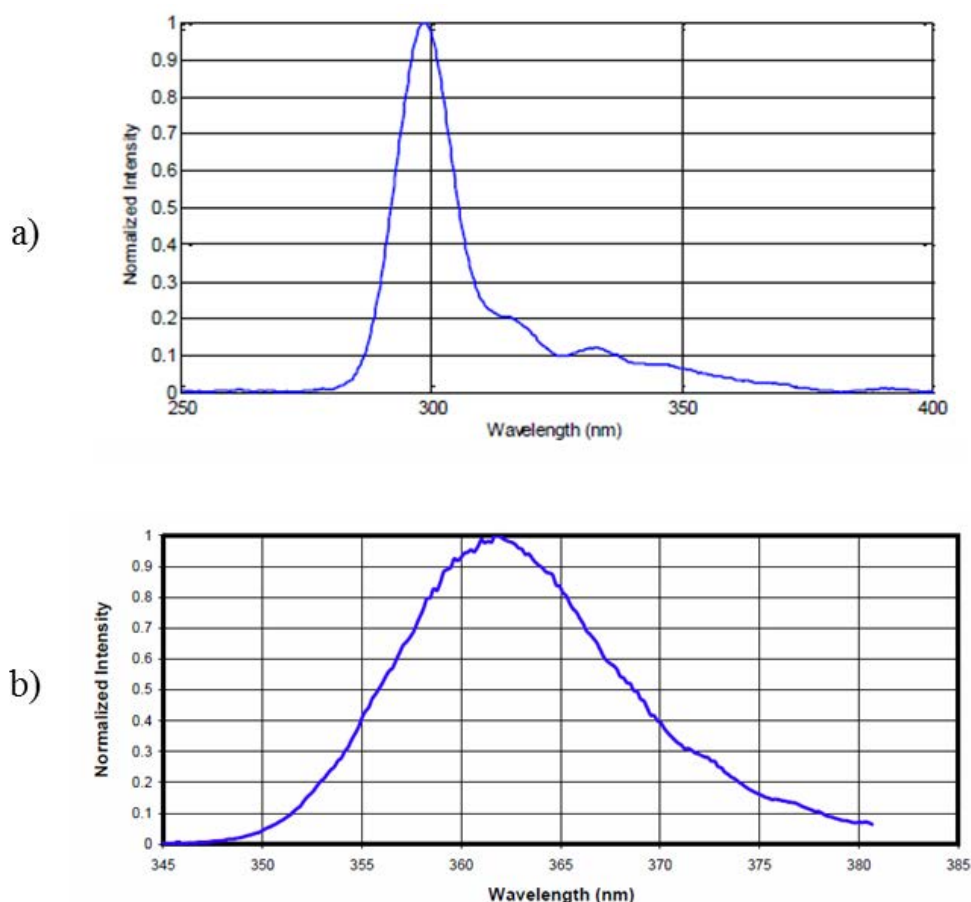


Figure 129. Spectres d'émission des diodes utilisées dans ce travail, d'après la documentation de Thorlabs ; a) LED300W ; b) LED365W.

Avec l'aide du Pr. Sophie Cribier au laboratoire, nous avons réalisé un montage facile à déplacer permettant d'irradier des échantillons. Les échantillons sont déposés dans une plaque 24 puits, placée sur de la glace. La diode électroluminescente est située à courte distance de l'échantillon (2 cm) et son faisceau est dirigé vers celui-ci, ce qui permet de contrôler la surface soumise au rayonnement.

Afin de tester l'efficacité de ces LEDs, les oligonucléotides 13-Bzp, 17-Iodo et 17-Thio ont été incubés en présence de hRPA dans les conditions permettant la formation des complexes. Ils ont ensuite été irradiés pendant 40 minutes avec une diode, placée à 2 cm de l'échantillon soit à une longueur d'onde de 300 nm soit à une longueur d'onde de 365 nm. En parallèle, les mêmes complexes hRPA-oligonucléotides ont été irradiés pendant 40 minutes avec la lampe UV classique. Les complexes ont ensuite été séparés sur un gel dénaturant comme dans les expériences précédentes afin d'évaluer le pourcentage de complexe covalent formé. Les résultats sont présentés dans la Figure 130.

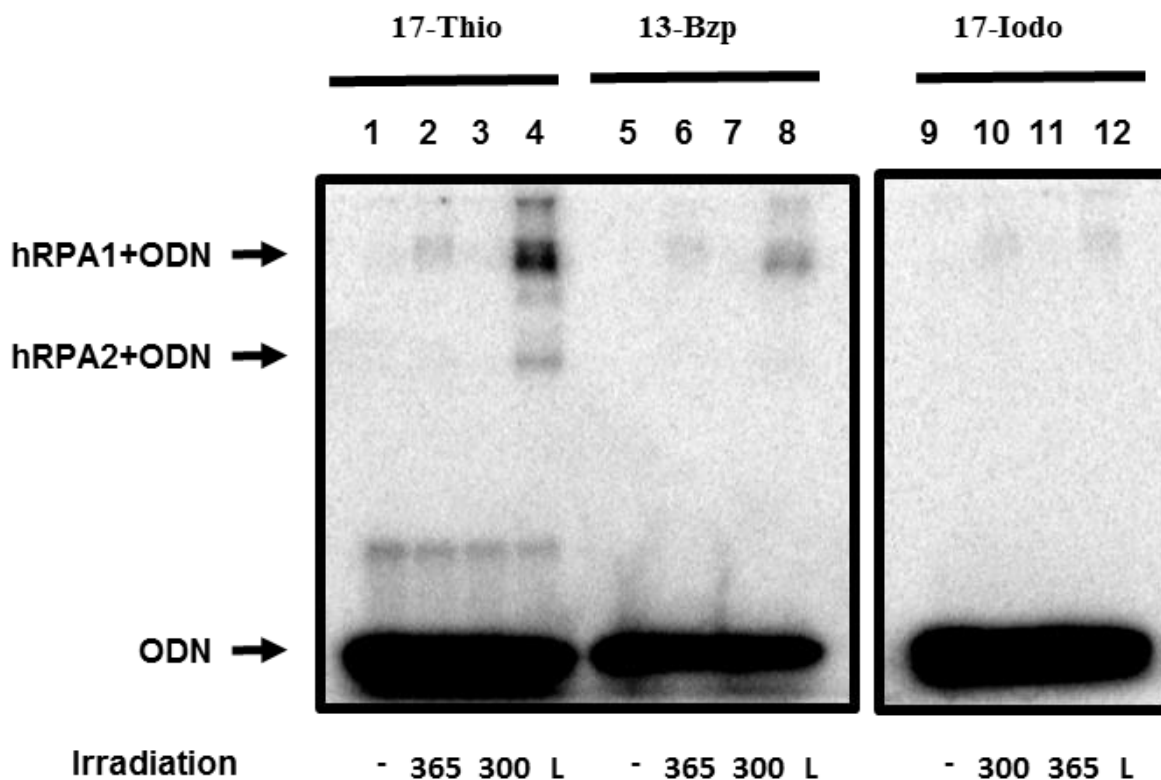


Figure 130. Utilisation des LEDS pour irradier les sondes 17-Thio, 13-Bzp et 17-Iodo en présence de hRPA (- : pas d'irradiation ; 365, 300 et L représentent respectivement les échantillons irradiés avec la LED à 365 nm, la LED à 300 nm et la lampe UV à 365 nm). La séparation a été effectuée sur un gel d'acrylamide dénaturant (Autoradiogramme).

L'irradiation du brin 17-Thio par la diode à 365 nm donne lieu à la formation de 0.5% d'adduits (piste 2), comparés à 11% avec la lampe classique (piste 4). Le rendement de photomarquage est donc nettement plus faible lorsqu'on utilise la diode.

L'irradiation du brin 13-Bzp par la diode à 365 nm donne lieu à la formation de 1% d'adduits (piste 6), comparés à 4.9% avec la lampe classique (piste 8). On constate donc que l'irradiation par le système de LEDs est moins efficace, pour ce temps d'irradiation, que l'irradiation par la lampe UV classique. L'irradiation de ces deux sondes par la LED à 300 nm ne permet pas d'observer la formation de d'adduits entre hRPA et les brins photoactivables (pistes 3 et 7). L'irradiation du brin 17-Iodo avec la LED à 300 nm ainsi qu'avec la lampe UV classique donne lieu à la formation de 0.4% d'adduits (pistes 10 et 12 respectivement). La LED à 365 nm ne permet pas d'observer la formation d'adduits (piste 11).

L'irradiation avec les LEDs a donc bien permis de former des adduits simple brin-RPA, de la même manière que lors de l'irradiation par la lampe UV préalablement utilisée, mais avec des rendements de photomarquage moins élevés qu'en utilisant celle-ci. Pour la

benzophénone et la ^{48}T , le gain que nous espérons obtenir du fait de la plage d'irradiation resserrée des LEDs est largement contrebalancé par la faible puissance de celles-ci (0.5 et 2 mW, comparé à 12W pour la lampe UV classique). Nous avons donc choisi de poursuivre nos expériences avec la lampe UV classique dans le cadre de cette étude. Dans le cas de la ^{51}dU , l'irradiation par la LED à 300 nm conduit au même rendement de photomarquage que l'irradiation avec la lampe UV ce qui, si l'on considère la différence de puissance des lampes, nous laisse penser que l'irradiation par la diode à 300 nm favorise donc bien le photomarquage par la ^{51}dU . Cependant, la faible efficacité de ce photomarquage nécessitera d'autres optimisations de ce protocole qui pourrait s'avérer intéressant pour irradier les sondes ^{51}dU sans avoir besoin d'un équipement coûteux comme un laser.

L'irradiation par les LEDs ne va pas ou peu réchauffer l'échantillon. On peut donc imaginer d'irradier les échantillons pendant plusieurs heures pour améliorer le rendement de photomarquage sans dégrader les protéines présentes dans l'échantillon. Cette augmentation devrait être transposable à toutes les sondes et plus particulièrement celles contenant une benzophénone, du fait de son mécanisme de régénération vu au début de ce chapitre. De même, on peut imaginer de rapprocher encore la LED de l'échantillon ce qui permettrait d'augmenter l'efficacité du photomarquage sans induire une augmentation drastique de la température. Enfin, l'utilisation d'une diode à 325 nm pourrait permettre l'amélioration du rendement de photomarquage de la ^{51}dU et de la ^{48}T .

2.5 Essais de photomarquage en présence d'un mélange de protéines

Nous avons trouvé des conditions de photomarquage efficaces pour la plupart de nos oligonucléotides en présence d'une protéine modèle purifiée et nous avons vu notamment que la diazirine possédait des propriétés intéressantes en terme de rendement et de cinétique. Afin de valider l'utilisation de notre sonde photoactivable la plus efficace sur un modèle plus complexe, c'est-à-dire un mélange de protéines, nous avons réalisé des expériences de photomarquage sur des extraits bactériens surexprimant ou non hRPA.

Des souches d'*E. coli* BL21 DE3 non transformées ou transformées avec un plasmide codant hRPA, ont été mises en culture et induites en présence d'IPTG. Après 2 heures

d'induction, des quantités équivalentes de culots de bactéries ont été lysés et la concentration en protéines totale des différents lysats a été évaluée par un test de Bradford. Des lysats cellulaires contenant des quantités équivalentes en protéines totales (1 mg/mL) ont alors été incubés pendant 30 minutes en présence de l'oligonucléotide 16-Dzr-Biot. Les échantillons ont été ensuite séparés en deux fractions : une fraction a été irradiée pendant 30 minutes avec la lampe UV classique, l'autre fraction n'a pas été irradiée. En parallèle, des échantillons contrôles ont été réalisés dans lesquels les oligonucléotides ont été incubés en présence de protéines purifiées.

Les échantillons ont été ensuite analysés sur un gel dénaturant qui est présenté dans la Figure 131.

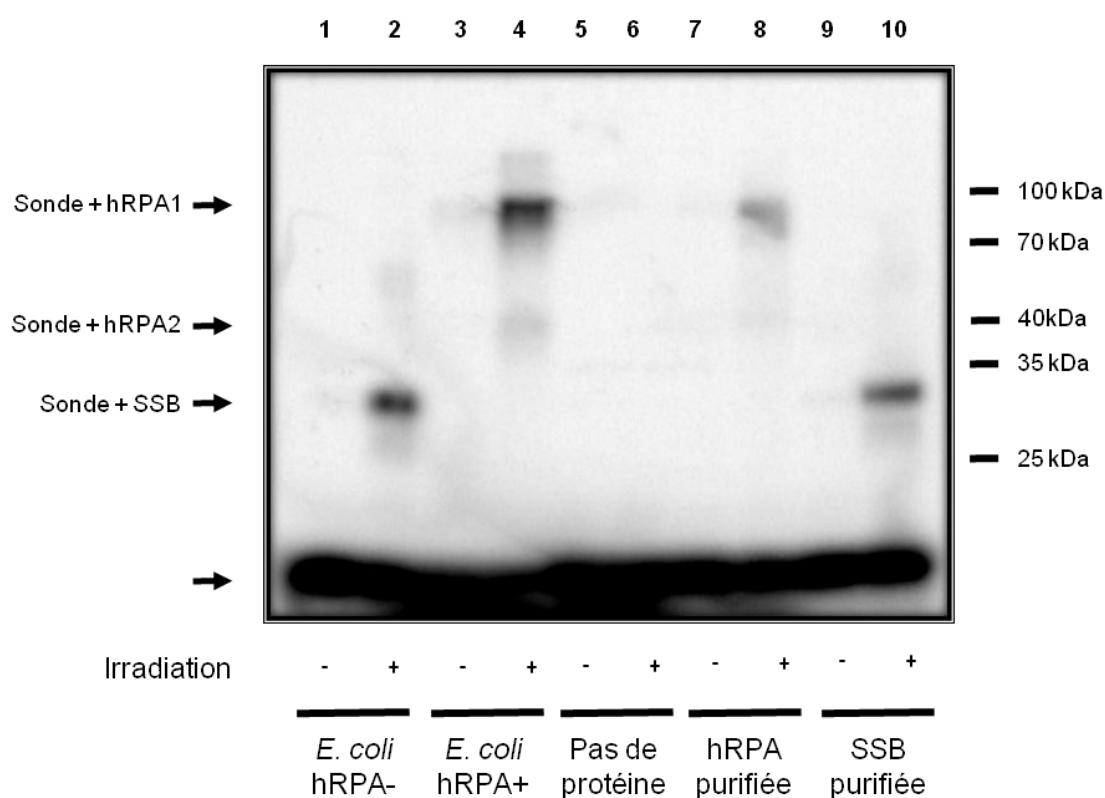


Figure 131. Irradiation d'extraits bactériens surexprimant ou non hRPA (Autoradiogramme). La séparation a été effectuée sur un gel d'acrylamide dénaturant.

Lorsque la sonde est irradiée en présence de lysats de bactéries ne surexprimant pas hRPA, on observe une bande très majoritaire qui correspond à une masse moléculaire un peu en dessous de 35 kDa (piste 2). Lorsque la sonde est irradiée en présence d'extraits de

bactéries surexprimant hRPA, cette bande disparaît totalement au profit d'autres bandes qui correspondent au photomarquage des sous-unités 1 et 2 de hRPA (piste 4) comme en témoigne notre contrôle en présence de hRPA purifiée (piste 8). Concernant la bande observée dans la souche ne surexprimant pas hRPA, nous avons émis l'hypothèse qu'il s'agissait de la protéine de liaison au simple brin (SSB) d'*E. coli*. Nous avons vérifié cette hypothèse en effectuant l'expérience de photomarquage en présence de SSB d'*E. coli* commerciale (piste 10) et nous avons constaté que la bande observée dans l'extrait hRPA- correspond effectivement à celle observée lors du photomarquage de la SSB commerciale.

Notre oligonucléotide comportant l'agent de photomarquage de type diazirine permet donc de photomarquer spécifiquement notre protéine modèle, hRPA, dans un milieu biologique complexe contenant de nombreuses protéines et biomolécules.

Afin d'évaluer la capacité de nos sondes à observer des différences d'affinité de liaison pour des protéines dans des extraits cellulaires, nous avons également effectué cette même expérience en présence d'un lysat bactérien d'une souche surexprimant un mutant de hRPA, hRPA D228Y. Cette mutation se situe dans la sous-unité hRPA1 et a pour conséquence de diminuer l'affinité de liaison de hRPA pour le simple brin (article en cours de soumission).

Ainsi, la sonde 16-Dzr-Biot a été incubée en présence de lysats cellulaires provenant de bactéries surexprimant soit hRPA sauvage (WT) soit le mutant hRPA D228Y (MUT). Les lysats provenant de bactéries non transformées (NT) ont également été testés comme contrôle.

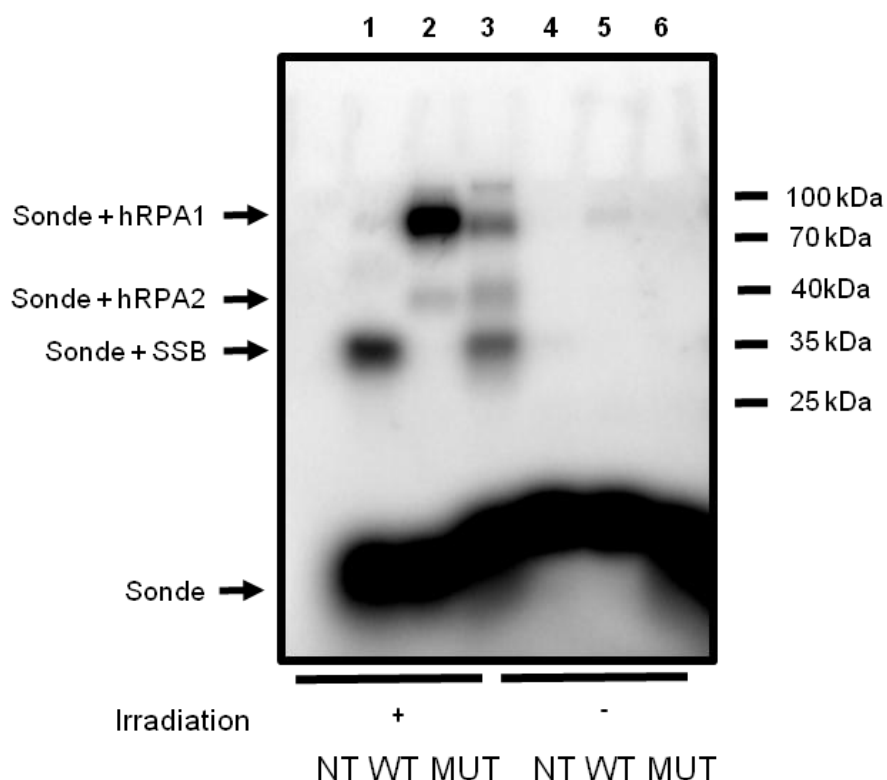


Figure 132. Irradiation de la sonde 16-Dzr-Biot en présence d'extraits bactériens surexprimant hRPA sauvage ou hRPA mutée (Autoradiogramme). La séparation a été effectuée sur un gel d'acrylamide dénaturant.

Les résultats obtenus après irradiation et séparation des échantillons sur un gel dénaturant sont présentés dans la figure 132. Lorsque la sonde est incubée et irradiée avec des lysats de bactéries non transformées (NT, piste 1), on observe de manière similaire à l'expérience présentée ci-dessus le photomarquage de la protéine de liaison au simple brin d'*E.coli*. De la même façon, ce photomarquage disparaît au profit de celui de hRPA dans les lysats surexprimant hRPA (WT, piste 2). Lorsque la sonde est incubée et irradiée avec les extraits surexprimant la hRPA D228Y, on observe d'une part que le photomarquage de la SSB est partiellement rétabli, en parallèle de celui de hRPA (piste 3). D'autre part, les rapports de photomarquage des deux sous-unités hRPA1/hRPA2 sont modifiés par rapport à ce que l'on observe avec les extraits comportant hRPA WT. Ainsi, il semble que la sous-unité hRPA1 soit moins photomarkuée dans l'échantillon D228Y avec un rapport hRPA1/hRPA2 de 3 : 1 alors qu'il est de 30 : 1 dans l'échantillon WT.

Ces observations sont bien en accord avec la diminution de l'affinité de hRPA D228Y pour le simple brin, qui entraîne la réapparition du photomarquage de SSB d'*E.coli*, et le fait

que cette mutation soit située dans hRPA1, qui entraîne la baisse de photomarquage de cette sous-unité comparativement à hRPA2.

Afin de vérifier que ces résultats n'étaient pas dus à des différences d'expression entre hRPA et hRPA D228A dans les différents lysats cellulaires, des westerns blots ont été réalisés sur les lysats bactériens avec des anticorps contre hRPA1 ou hRPA2. On observe que hRPA non mutée est exprimée en plus grande quantité (3 à 4 fois plus) que hRPA D228Y, mais dans des proportions qui ne peuvent pas expliquer entièrement la réapparition du photomarquage de la SSB (Figure 132).

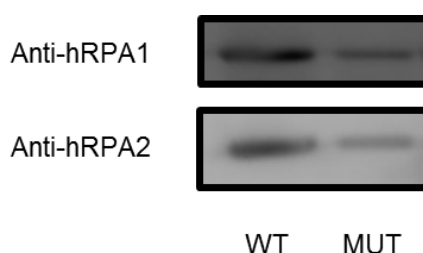


Figure 133. Western blot des lysats bactériens surexprimant la protéine hRPA (WT) ou le mutant hRPA D228Y (MUT). La quantité de protéines totales des différents extraits a été dosée et la même concentration de protéines a été déposée dans chacun des puits.

La sonde diazirine est donc bien adaptée à l'étude de milieux complexes. Elle permet de réaliser du photomarquage sur une protéine modèle, que ce soit dans un modèle simple ou sur des extraits cellulaires. De plus, nous nous sommes aperçus que cette sonde pouvait être utilisée de façon pertinente pour étudier la protéine hRPA. En effet, l'étude du photomarquage de la sonde diazirine avec des mutants de hRPA a permis de mettre à jour une différence d'affinité pour l'ADN entre une sous-unité de la protéine sauvage et mutante. L'utilisation de nos sondes photoactivables pourrait donc permettre d'obtenir des informations intéressantes sur les interactions entre hRPA et ADN.

2.6 Mise au point des conditions de capture sur bille de streptavidine

A terme, nos sondes oligonucléotidiques ont été conçues pour du photomarquage sur extraits cellulaires en vue d'identification de protéines par spectrométrie de masse. Du fait d'un rendement parfois faible de la réaction de photomarquage ou de la faible affinité de certains partenaires protéiques pour l'ADN, un enrichissement à l'aide du système

biotine/streptavidine est généralement indispensable pour permettre d'accéder aux protéines associées. C'est pourquoi nous avons placé une biotine sur le brin comportant l'agent de photomarquage. Il était donc indispensable de mettre au point, d'une part les conditions de capture sur des billes magnétiques fonctionnalisées avec de la streptavidine et, d'autre part, les conditions de relargage.

Nous avons utilisé les billes Dynabeads® M-280 Streptavidin d'Invitrogen très utilisées au laboratoire et suivi un protocole de lavage adapté d'une publication du laboratoire ²⁶². Après incubation des oligonucléotides (simple brin ou duplex, marqués sur le brin contenant la biotine) avec les billes magnétiques pendant une heure à température ambiante, les billes sont retenues par un support magnétique puis lavées successivement avec les tampons A, B et C présentés ci-dessous, et finalement avec de l'eau bidistillée.

Tampon A 50 mM Tris pH=7.4 ; 0.1 mg/mL BSA

Tampon B 50 mM Tris pH=7.4 ; 0.1 mg/mL BSA ; 0.1% SDS

Tampon C 50 mM Tris pH=7.4 ; 0.1 mg/mL BSA ; NaCl 1M

Afin d'évaluer l'efficacité de notre protocole de capture, nous avons suivi à chaque étape, la présence ou l'absence des oligonucléotides en estimant la radioactivité dans les solutions de lavage avec un compteur à scintillation. Ainsi, nous avons pu observer que les oligonucléotides simple brin contenant une fonction biotine à leur extrémité 3' étaient bien retenus sur bille tandis que des contrôles réalisés en parallèle ne possédant pas de biotine étaient éliminés par lavage (0.9% de la radioactivité initiale restante sur les billes après la série de lavages). La capture sur bille est bien biotine-dépendante et ne se fait pas par des interactions non spécifiques.

La mise au point des conditions de relargage des duplex a été plus délicate et nous avons testé différentes conditions qui sont récapitulées dans le Tableau 8. L'incubation à 95°C pendant 15 minutes soit dans une solution d'eau bidistillée, une solution aqueuse d'HCl 0.1M, une solution aqueuse d'EDTA 10 mM ou une solution aqueuse de SDS 0.1% n'a pas eu pour conséquence de séparer significativement les oligonucléotides des billes. En revanche, l'incubation à 95°C dans du tampon de charge d'électrophorèse (entrée 5) a permis de rompre la liaison biotine-streptavidine.

Entrée	Condition	% relargage (estimation par comptage par scintillation)
1	Eau bidistillée, 15 min, 95 °C	10 %
2	HCl 0.1M, 15 min, 95 °C	3 %
3	EDTA 10 mM, 15 min, 95 °C	8 %
4	0.1% SDS, 15 min, 95 °C	16 %
5	50 mM Tris pH=6.8 ; 8% glycerol ; 1.25% SDS ; 0.02% BBP, 0.1 mM β-mercaptoéthanol, 15 min, 95 °C	86 %

Tableau 8. Détermination des conditions de relargage des oligonucléotides.

Après avoir trouvé des conditions de capture des oligonucléotides sur les billes magnétique, de lavages ainsi que de relargage, nous avons vérifié que nous étions capables de réaliser cette succession d'étapes lorsqu'une protéine était photomarkée par nos oligonucléotides. A cet effet, des échantillons contenant la sonde 1-Thio-Biot ou la sonde 1-Thio, deux séquences ne différant que par l'ajout d'une fonction biotine, ont été incubés avec hRPA pendant 20 minutes à température ambiante. Les échantillons ont été ensuite irradiés pendant une heure avec la lampe à 365 nm (ou conservés à l'abri de la lumière pour les contrôles non irradiés). Chacun des échantillons a été ensuite séparé en deux. Une partie a été soumise à une capture/lavage/relargage avec les billes M-280, tandis que l'autre a été déposée directement sur le gel dénaturant d'acrylamide.

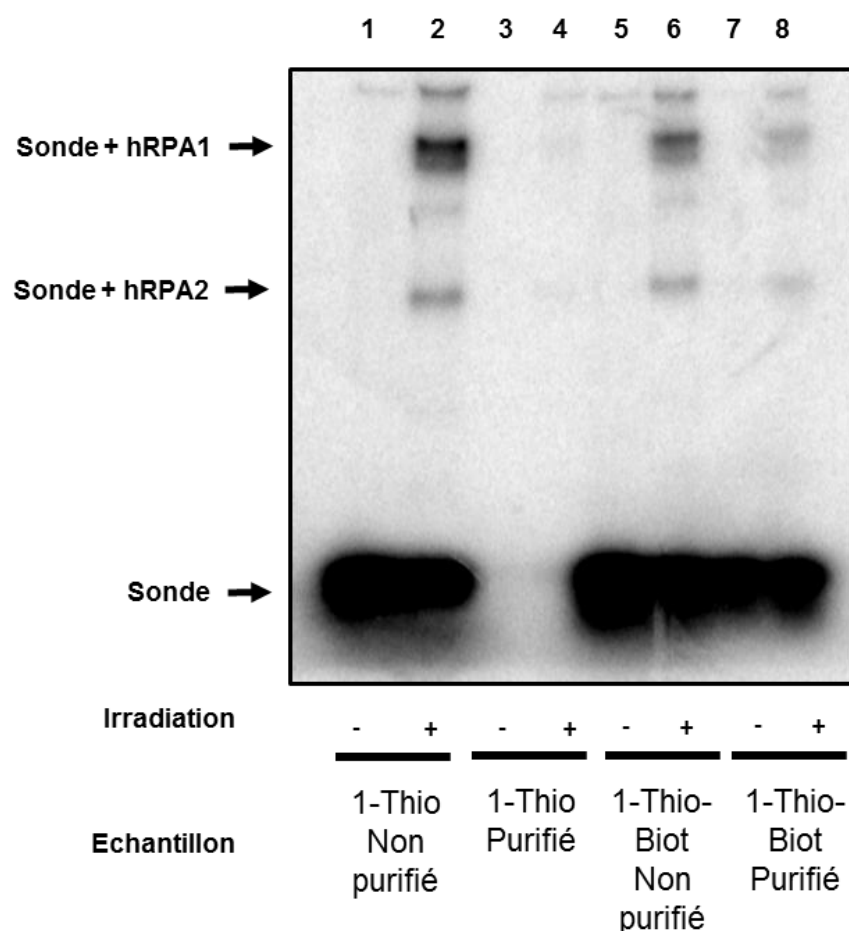


Figure 134. Photomarquage, capture et relargage de hRPA en présence d'une sonde avec (1-Thio-Biot) ou sans (1-Thio) biotine (Autoradiogramme).

Le gel obtenu est présenté dans la figure 134. Dans le cas des échantillons directement déposés sur le gel, on observe le photomarquage de hRPA1 et hRPA2 pour les deux sondes (pistes 2 et 6). Dans le cas des échantillons purifiés à l'aide des billes paramagnétiques, uniquement ceux avec la sonde 1-Thio-Biot, qui contient une biotine, font apparaitre des complexes covalents entre hRPA et la sonde (piste 8). Les échantillons purifiés de 1-Thio (qui ne contient pas de biotine) ne présentent aucune radioactivité, que ce soit au niveau des complexes hRPA-oligonucléotide ou bien de l'oligonucléotide seul. En effet, les adduits covalents formés ainsi que l'oligonucléotide simple brin non complexés ont été éliminés des billes par les lavages successifs. Nous avons donc mis en place les conditions permettant de photomarquer la protéine hRPA avec un oligonucléotide photoactivable, de capturer le complexe covalent formé et de le relâcher en utilisant des billes magnétiques.

Cependant, la comparaison entre les pistes 6 et 8 nous permet de constater que ce protocole de capture/relarguage induit une perte de signal. La quantification des niveaux de radioactivité nous permet d'estimer qu'entre 23 et 39% des oligonucléotides présents à l'origine sont détectés après le relarguage des billes magnétiques. Cette perte de signal est compensée par le fait que l'utilisation des billes magnétiques permet de concentrer les sondes oligonucléotidiques présentes dans des échantillons de grand volume. Une autre façon d'y remédier est de lier les séquences d'ADN sur les billes de streptavidine avant l'incubation avec les extraits protéiques et l'irradiation. En effet, l'une des origines de cette perte de signal est que la fonction biotine est moins accessible après que la sonde oligonucléotidique ait formé une liaison covalente avec des protéines.

2.7 Essais de photomarquage sur des extraits nucléaires

Nos sondes ont été conçues pour tenter de caractériser des protéines susceptibles de reconnaître spécifiquement le motif 5hmC au sein d'une séquence d'ADN. Nous avons donc réalisé quelques essais préliminaires de photomarquage avec le duplex comportant la sonde diazirine en présence d'extraits nucléaires commerciaux de cellules HeLa, ainsi que des extraits nucléaires de cerveau de rat. En comparant les profils de photomarquage sur gel d'électrophorèse lorsque les extraits sont incubés en présence du duplex ne comportant pas la modification épigénétique (Cds16-Dzr-Biot) et le duplex comportant la 5hmC (Hds16Dzr-Biot), nous cherchions à mettre en évidence des différences entre les deux profils et ainsi déterminer si des protéines présentes dans ces extraits interagissaient spécifiquement avec la 5hmC.

Nous avons d'abord fait un premier essai avec des extraits commerciaux de cellules HeLa. Les sondes Cds16Dzr-Biot et Hds16Dzr-Biot ont été incubées à une concentration de 5 nM en avec les extraits nucléaires (2mg/mL quantifiés par dosage au réactif de Bradford) ou bien dans le tampon d'incubation. Après incubation des duplex pendant 30 minutes à 4°C en présence ou non d'extraits nucléaires, puis irradiation pendant 30 minutes à 365 nm, les échantillons ont été déposés sur un gel dénaturant d'acrylamide (Figure 135).

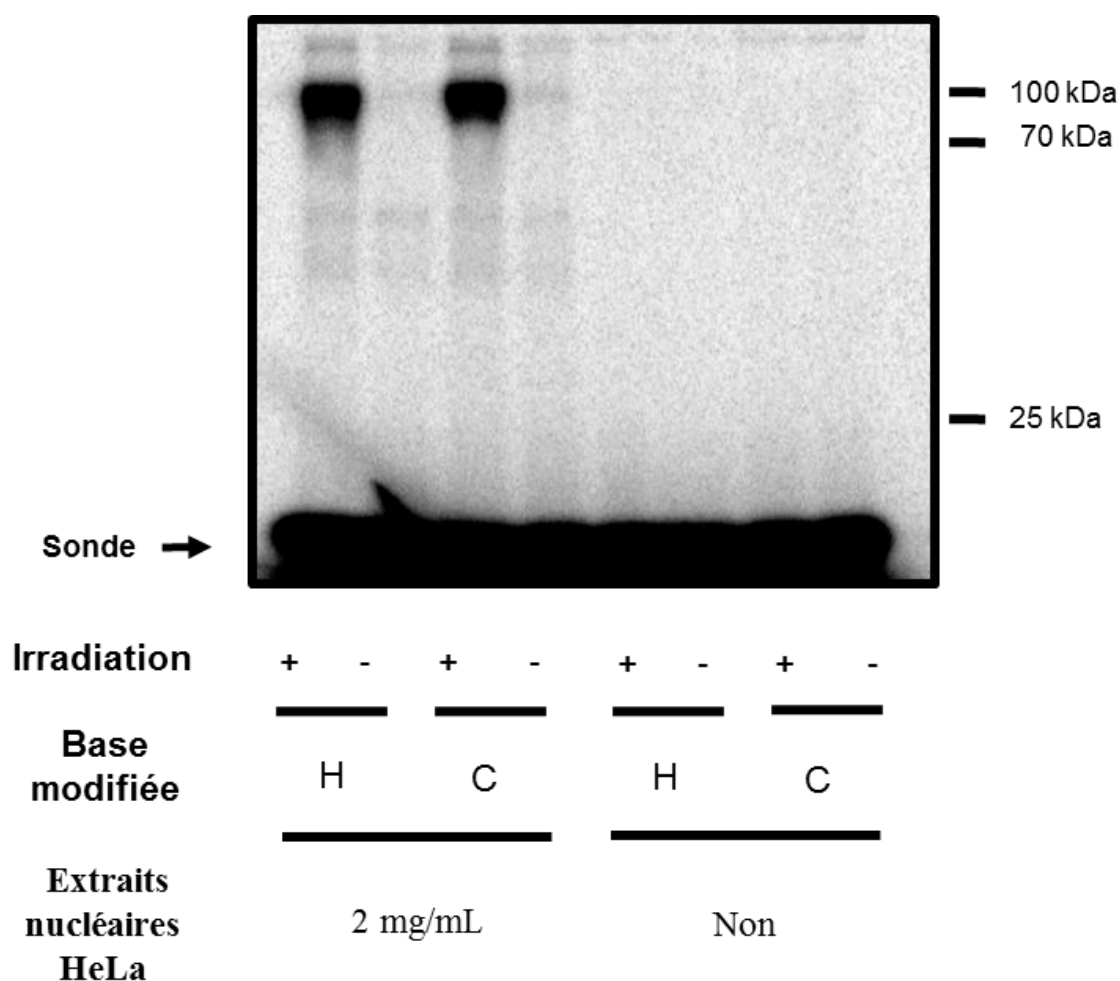


Figure 135. Photomarquage d'extraits nucléaires de cellules HeLa par les sonde Cds-16-Dzr-Biot et Hds-16-Dzr-Biot et séparation sur gel dénaturant d'acrylamide (autoradiogramme).

L'analyse du gel dénaturant nous permet d'identifier une bande préférentiellement photomarkée à une masse moléculaire aux alentours de 100 kDa. Cette bande est présente pour la sonde Cds-16-Dzr-Biot ainsi que pour la sonde Hds-16-Dzr-Biot. Le gel ne montre donc pas de différence notable de profil de photomarquage entre les deux duplexes. Même si ce résultat préliminaire est décevant, cette absence de différence apparente est à relativiser : en effet, si nous espérons observer des bandes différentes, l'analyse par gel d'acrylamide ne permet pas d'identifier une protéine photomarkée sur le gel mais uniquement de détecter sa présence et sa masse apparente. Deux protéines différentes peuvent ainsi migrer à la même masse et une bande peut en fait contenir plusieurs protéines différentes.

Nous avons réalisé ces essais préliminaires avec des extraits protéiques facilement accessibles issus de cellules HeLa qui sont des cellules cancéreuses. Comme on l'a vu

dans l'introduction générale, elles ne contiennent en fait que de faibles niveaux de 5hmC et ne constituent donc pas forcément le meilleur modèle pour caractériser de nouvelles protéines interagissant avec la 5hmC. Les tissus les plus riches en 5hmC proviennent, comme nous l'avons vu dans l'introduction, du cerveau et il nous a donc semblé indispensable de travailler sur des extraits protéiques issus de ces tissus. Avec l'aide de Frédéric Thiebaut, qui vient de commencer une thèse au laboratoire, nous avons donc réalisé un autre essai de photomarquage sur des extraits nucléaires de cerveau de rats. Les extraits nucléaires ont été préparés à partir de cerveaux entiers avec le kit CellLytic™ NuCLEAR™ Extraction, de Sigma-Aldrich. Les sondes Cds-16-Dzr-Biot (C), Hds-16-Dzr-Biot (H) et 16-Dzr-Biot (SS) ont ensuite été incubées pendant 40 minutes en présence de différentes quantités d'extraits nucléaires (mesurées par un test de Bradford), puis les échantillons ont été déposés sur un gel dénaturant d'acrylamide et séparés (Figure 136).

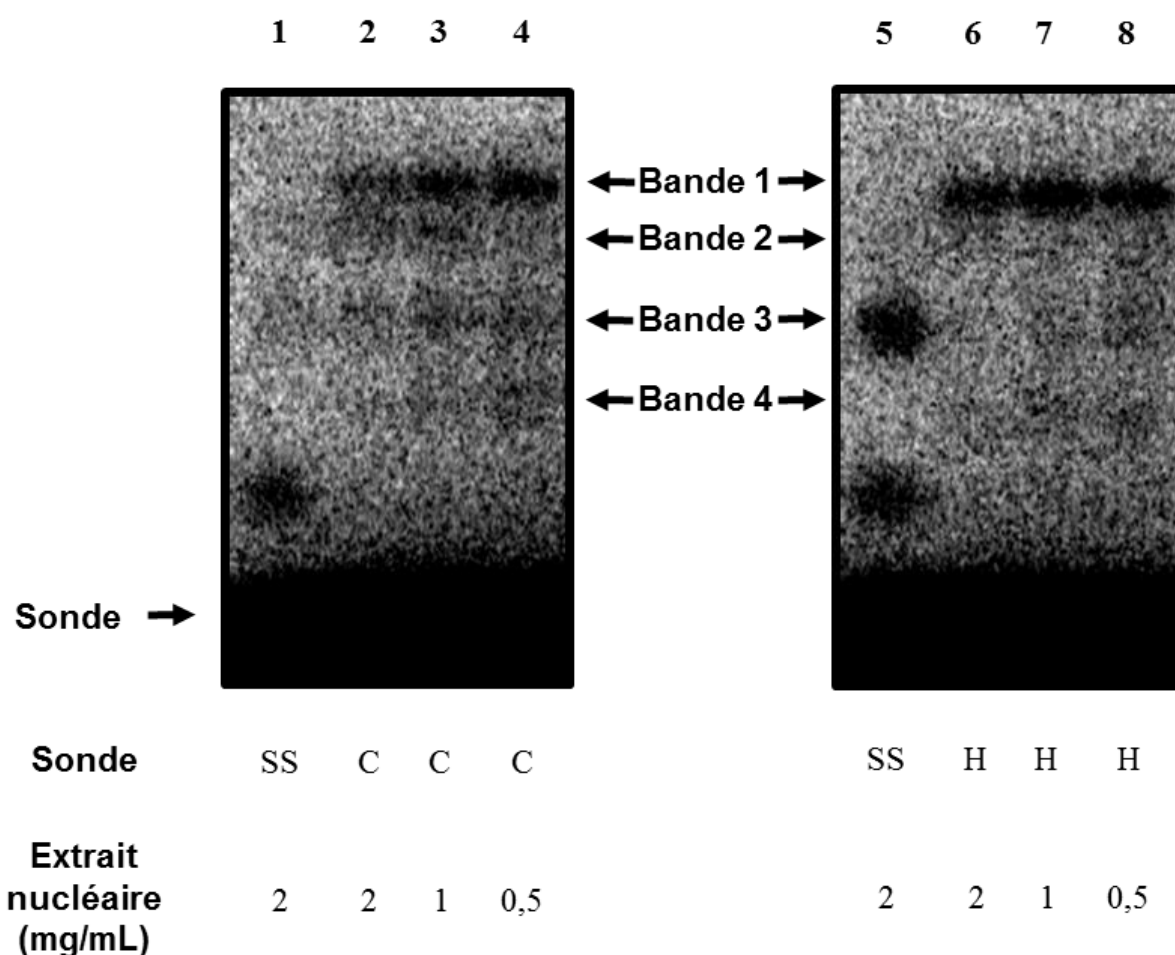


Figure 136. Photomarquage d'extraits nucléaires de cerveau de rats par les sondes Cds-16-Dzr-Biot (C), Hds-16-Dzr-Biot (H) et 16-Dzr-Biot (SS).

On observe à nouveau le photomarquage d'une bande majoritaire (Bande 1), quelles que soient les quantités d'extraits cellulaires ajoutées dans l'échantillon, et la nature de la sonde duplex. Cependant, cette fois-ci, on peut apercevoir des bandes faiblement présentes à des masses moins importantes que celle de la bande majoritaire (Bandes 2, 3 et 4). En particulier, la bande 2 est présente pour la sonde contenant une cytosine et pas pour celle contenant une 5hmC. Notre expérience de photomarquage sur des extraits protéiques nucléaires de cerveau de rats semble donc faire apparaître une différence dans les profils de protéines photomarkées lors de l'analyse des gels radioactifs mais il est clair que cette expérience devra être répétée et notamment avec des sondes plus fortement radiomarkées afin d'avoir une vision qualitative et quantitative plus précise de ces différences.

3 Conclusion et perspectives

Dans ce chapitre, nous avons donc présenté la synthèse et l'évaluation de sondes oligonucléotidiques pour des applications de photomarquage d'affinité. Nous avons conçu ces sondes pour permettre de capturer et d'identifier des protéines interagissant spécifiquement avec la 5-hydroxyméthylcytosine. Pour cela, nous avons choisi de placer la base modifiée de l'ADN sur un brin d'un duplex de l'ADN et d'introduire un groupement photoactivable ainsi qu'une fonction biotine sur le brin complémentaire. Afin de permettre une modulation simple et efficace des groupements photoactivables, nous avons choisi d'introduire ceux-ci à l'aide d'une méthode de fonctionnalisation post-synthétique, dite du « convertible nucleoside ». Grâce à cette méthode, nous avons obtenu des séquences d'ADN comportant une benzophénone ou une trifluorométhyl diazirine. Des séquences comportant d'autres groupements photoactivables, la 5-Iodo-2'-désoxyuridine et la 4-Thio-2'-désoxythymidine ont aussi été obtenues par synthèse sur support solide. Nous avons vérifié que l'introduction des différentes modifications dans nos sondes photoactivables ne déstabilisait pas les duplex correspondants.

Les propriétés de ces quatre groupements photoactivables dans le contexte de sondes d'ADN simple brin ont ensuite été évaluées à l'aide d'une protéine modèle. L'étude de l'interaction et du photomarquage de la hRPA, protéine de liaison au simple brin humaine, a

permis de mettre en évidence des cinétiques de photomarquage différentes. Nous avons aussi montré qu'il était possible de photomarquer spécifiquement hRPA dans un milieu complexe, et de capturer, purifier et relarguer les complexes covalents ADN-hRPA formés lors de l'irradiation. Le photomarquage d'un mutant de hRPA a de plus permis de montrer que nous pouvions faire apparaître des différences d'affinité lors du photomarquage. Des expériences de complexation et de photomarquage avec la protéine Dnmt3a nous ont permis de valider nos sondes sur des protéines reconnaissant l'ADN double brin.

Au final, nous avons mis au point les outils et les techniques pour l'utilisation de ces sondes que nous avons tenté de mettre en application à la fin de la thèse. Le but final de nos sondes était de permettre la caractérisation de nouvelles protéines interagissant avec la 5hmC. Nous avons donc réalisé quelques essais de photomarquage sur des extraits cellulaires, suivis d'une analyse par gel d'acrylamide, pour essayer d'observer des différences dans les profils de protéines photomarkées selon la présence ou l'absence de 5hmC dans les sondes utilisées. Les premiers essais d'utilisation de nos sondes sur des extraits nucléaires de cellules HeLa n'ont pour l'instant pas permis de faire apparaître des différences de photomarquage entre la sonde comportant une cytosine non modifiée et celle comportant une 5-hydroxyméthylcytosine. Utilisées en présence d'extraits nucléaires de cerveau de rats, nos sondes ont fait apparaître des différences de profil qu'il faudra comprendre et exploiter. Il sera par exemple intéressant de reproduire ces expériences avec les différentes sondes que j'ai mentionnées mais également avec d'autres que j'ai synthétisées mais je n'ai pas évoquées car pas utilisées dans les travaux présentés. J'avais en effet lors des campagnes de synthèse d'oligonucléotides conçu également des sondes permettant l'introduction du groupement photoactivable à différentes positions. Il pourra aussi être intéressant également de travailler sur des extraits nucléaires de différentes parties du cerveau pour voir s'il existe des différences dans la mesure où des différences de niveaux de 5hmC ont été détectées selon la zone du cerveau⁴⁶.

Afin d'exploiter au mieux ces sondes, l'étape suivante consistera à utiliser des techniques de protéomique et d'analyse par spectrométrie de masse. La combinaison de photomarquage d'extraits cellulaires suivie d'une purification par billes de streptavidine nous permettra de ne récupérer que les protéines étant liées aux sondes utilisées. Une digestion protéique suivie d'une analyse par spectrométrie de masse nous permettra d'accéder aux différents profils de protéines photomarkées et d'identifier de nouvelles protéines reconnaissant la 5hmC.

Très récemment, un exemple d'une telle méthode appliquée à l'étude de la 5hmC a été présenté dans l'étude du groupe de Davis, abordée dans l'introduction de ce chapitre²⁵⁶. Ce groupe a entre autre synthétisé un brin d'ADN photoactivable contenant une diazirine et une biotine. L'ajout par PCR d'un brin complémentaire comportant soit une 5hmC soit une cytosine non modifiée a permis de former les sondes photoactivables finales. Ces sondes ont ensuite été utilisées, comme dans nos premiers essais, pour photomarquer des extraits nucléaires de cellules HeLa. La comparaison des différentes protéines identifiées par spectrométrie de masse en fonction de la nature de la sonde a permis d'identifier des protéines de liaison à l'ADN interagissant préférentiellement avec la 5-hydroxyméthylcytosine. De manière similaire, l'utilisation de nos différentes sondes nous permettra d'obtenir des informations différentes sur les protéines interagissant avec la 5hmC.

Conclusion générale

Mes travaux de thèse se sont articulés autour de deux axes : (i) Le développement de méthodes pour la quantification, la détection ou l'enrichissement de la 5hmC dans des échantillons d'ADN génomique et (ii) la synthèse et l'évaluation de sondes oligonucléotidiques photoactivables destinées à l'identification de nouvelles protéines interagissant avec la 5hmC.

J'ai d'abord adapté une technique de quantification globale de la méthylation de l'ADN par LC-MS/MS à l'étude de l'hydroxyméthylation de l'ADN. Après en avoir vérifié la sensibilité et la reproductibilité, j'ai appliqué cette technique d'analyse à des échantillons d'ADN extraits de cellules neuronales et embryonnaires de souris, dans lesquelles les premières identifications de la 5hmC ont eu lieu. La 5hmC a été détectée dans les cellules neuronales, mais pas dans les cellules souches embryonnaires. Appliquée à différentes lignées cancéreuses, la méthode n'a permis de détecter la 5hmC qu'en de rares occasions. Dans notre configuration, cette méthode ne semble donc pas adaptée à la quantification fine de la 5hmC, en raison de sa faible présence dans l'ADN. En effet, il sera difficile de faire des analyses quantitatives comparatives lorsque l'on se situe à la limite de détection, compliquant l'utilisation d'une technique déjà lourde et coûteuse. Cette technique reste cependant très intéressante pour effectuer la quantification globale de la méthylation comme nous l'avons vu dans le cadre d'une collaboration avec Joëlle Riond et Paola Arimondo à l'ETaC. Les résultats d'une nouvelle méthode d'analyse indirecte, alliant détection simultanée de la méthylation et du cycle cellulaire, ont pu être validés par nos analyses par spectrométrie de masse. Ces travaux ont donné lieu à publication dans le journal *Epigenetics*.

Dans un domaine de recherche très compétitif, nous avons réalisé des avancées dans le développement de méthodes de marquage spécifique de l'ADN : une stratégie enzymatique, se basant sur une glucosyltransférase du phage T4, ainsi qu'une stratégie chimique, se basant sur l'oxydation sélective de l'alcool de la 5hmC. Pour la méthode de marquage enzymatique, nous avons réalisé la synthèse de l'UDP-6-N₃-glucose, un substrat modifié donneur de glucose comportant une fonction azoture. Afin d'optimiser le transfert de ce glucose modifié sur la 5hmC, nous avons obtenu un mutant de l'enzyme BGT. Cependant, nous n'avons pas réussi à caractériser le transfert de ce substrat modifié sur notre duplex d'ADN modèle. Cela peut être dû au fait que notre séquence modèle ne soit pas reconnue par la BGT, et nous avons récemment obtenu de nouvelles séquences, comportant notamment un plus grand nombre de

5hmC, afin de vérifier cette hypothèse. La méthode de marquage chimique a été mise au point sur un nucléoside modèle, et nous avons pu caractériser l'oxydation de l'alcool allylique de la 5hmdC et la fonctionnalisation de la 5fdC par une biotine-hydrazine. Lors de l'application de la méthode sur un oligonucléotide modèle, nous avons été confronté à des problèmes de détection qui ne nous ont pas permis de caractériser sa fonctionnalisation. Il faudra probablement essayer d'autres conditions afin de mettre en évidence cette fonctionnalisation et, par exemple, l'utilisation d'agents fluorescents reliés à une hydrazine pourra s'avérer utile pour la détection. Par ailleurs, notre travail pourra trouver une autre voie d'application, qui serait d'améliorer la détection de la 5hmC dans la technique de quantification globale par spectrométrie de masse évoquée plus haut. En effet, ces deux étapes de fonctionnalisation pourraient être appliquées au mélange de nucléosides obtenus après digestion de l'échantillon d'ADN afin d'introduire sur la 5hmC un groupement permettant d'augmenter son ionisation lors de l'analyse MS/MS.

Enfin, j'ai synthétisé des oligonucléotides d'ADN contenant des nucléosides photoactivables à l'aide d'une méthode de fonctionnalisation post-synthétique. Cette méthode dite du « convertible nucleoside » a permis l'introduction d'une benzophénone et d'une trifluorométhyl diazirine dans les séquences d'ADN désirées. Les propriétés de photomarquage de ces groupements photoactivables ont ensuite été comparées à celle de nucléosides commerciaux, la 5-Iododésoxyuridine et la 4-Thiothymidine, à l'aide de protéines modèles : la protéine de réplication A humaine (hRPA), et la Dnmt3a. Des expériences de cinétique de photomarquage ont permis de faire une étude comparative des différentes propriétés photochimiques des groupements photoactivables. De plus, les expériences du photomarquage de la protéine hRPA et de son mutant D228Y ont mis en exergue l'utilité de ces sondes pour mettre en évidence les différences dans l'affinité pour l'ADN. Ces résultats seront présentés dans une publication qui est en cours de rédaction.

Nous avons également montré que des duplex comportant ces sondes photoactivables ainsi qu'une biotine sur un brin et une modification épigénétique sur le brin complémentaire, pourraient être utilisés pour l'identification de protéines partenaires des 5hmc. En effet, nous avons vu que (i) ces duplex ne sont pas déstabilisés malgré les nombreuses modifications introduites ; (ii) les conditions de photomarquage ont été optimisées et la sonde diazirine semble la plus appropriée pour le piégeage de protéines modèles ; (iii) les conditions d'enrichissement sur billes de streptavidine des complexes photomarqués ADN-protéines sont

maintenant au point ; (iv) des essais préliminaires sur des extraits nucléaires ont été réalisés et nous avons pu observer dans ces mélanges complexes que nos sondes permettaient de photomarquer des protéines qu'il restera à identifier. Nous avons donc mis en place tous les outils et les conditions de leur utilisation pour effectuer des études poussées de protéomique qui permettront de faire apparaître les différents profils protéiques photomarkés par les sondes. Ce travail est en cours, puisqu'il fait en partie l'objet de la thèse de Frédéric Thiébaud en collaboration avec Carole Saintomé et Emmanuelle Sachon. L'étude protéomique est en effet en train d'être mise en place à la Plateforme de protéomique de l'UPMC et le modèle qui sera probablement conservé est celui des extraits nucléaires de cerveau de rat. Il pourra être ainsi intéressant de comparer les profils protéiques aux taux d'hydroxyméthylation trouvés dans l'ADN génomique d'échantillons extraits de différentes zones du cerveau. Parmi les autres perspectives de ce travail, la méthodologie de synthèse et de photomarquage étant maintenant bien en place au laboratoire, elle pourra permettre d'effectuer des études sur d'autres bases modifiées de l'ADN, comme la 5fC, la 5caC en encore la base J des trypanosomes.

Partie expérimentale

Généralités

Materials. Solvents and most of the starting materials were purchased from Sigma Aldrich, TCI, VWR, Merck and Carlos Erba. All solvents were of analytical grade.

Methods. ^1H and ^{13}C NMR spectra were recorded on a Bruker Avance III spectrometer, operating at a ^1H frequency of 500.11 MHz, or on a Bruker DRX 400, operating at a ^1H frequency of 400 MHz. Chemical shifts are given in parts per million (ppm) using solvent residual signals as the reference and coupling constants (J) are given in Hertz (Hz). All yields reported are not optimized. Flash chromatography were performed using silica gel Merck 60 with 0.040–0.063 μm . HPLC purifications were performed on a Waters HPLC apparatus (Waters™ 1525 Binary HPLC Pump with Waters™ 2487 Dual λ Absorbance Detector) using a Xterra Prep RP18 column (7.8*300 mm, 10 μm) and purity was checked using a Waters Xterra MS C18 column (4.6*250 mm, 5 μm). Absorbance spectrophotometry was performed on a Nanodrop-2000c spectrophotometer (Thermo Scientist). High Resolution Mass Spectrum (HRMS) were recorded at the Institut Parisien de Chimie Moléculaire (FR 2769) of UPMC (electrospray source).

Partie expérimentale du chapitre II

1. Genomic DNA purification and hydrolysis

Genomic DNA was prepared using the Blood & Cell Culture DNA Mini Kit (Qiagen, Courtaboeuf, France) and quantified by the NanoDrop technology (ThermoScientific). DNA hydrolysis was performed as previously described¹⁴². Briefly, 2 µg of DNA were digested by Nuclease P1 (2h at 45°C in 0.01M AcONH₄ pH=5,3 buffer), snake venom phosphodiesterase (2h at 37°C in 0.1M NH₄HCO₃ buffer) and antarctic alkaline phosphatase (1h, 37°C).

2. LC-ESI mass spectrometry

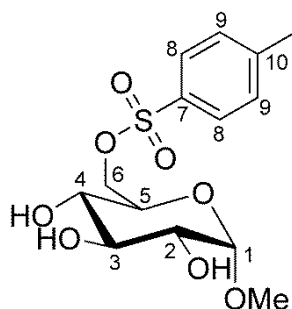
Genomic DNA methylation analysis was performed as previously described¹⁴². Nucleotides chromatography and mass spectrometry analysis were performed using a DionexUltimate 3000 HPLC (LG3400 micropump and WPS3000 autosampler) coupled to an Applied Biosystems QTrap mass spectrometer through a TurboIonSpray ion source interface. A sample equivalent to 1 µg of digested DNA (10 µl final volume) was injected in a Synergi 4u Polar-RP 80A 1.0 X 150 mm column (Phenomenex) and eluted with buffer A (0.1% formic acid in water) and B (0.1% formic acid in methanol) using a gradient increase of 0 to 50% B in 30 minutes.

The ESI conditions were as follows: nebulizer gas flow GS1= 20, curtain gas flow CUR= 20, collision gas flow CAD= high, Ion Spray voltage IS = +5500V, Declustering potential DP= 20V, Entrance potential EP= 10V, collision energy CE= 10V; collision cell exit potential CXP= 15V. LC-ESI-MS/MS chromatograms were acquired in Multiple Reaction Monitoring (MRM) mode by monitoring 4 transition pairs of molecular/fragment positive Ion at m/z 258.1/142.1, 242.1/126.1, 228.2/112.1, 268.1/152.1, for 5hmdC, 5mdC, dC, dG, respectively. The scan time was set to 200 ms for each ion pair. Cytosine methylation levels were calculated using the following formula: MeC (%) = $A_{5mdC} / (100 * (A_{5mdC} + A_{dC}))$ where A_{5mdC} and A_{dC} are respectively the integrals of the MRM peaks corresponding to 5mdC and dC.

Partie expérimentale du chapitre III.

1. Synthèse de l'UDP-6-N₃-glucose

6-*O*-toluenesulfonyl- α -methyl-D-glucopyranoside (**2**)



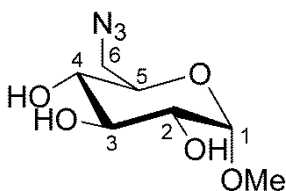
To a solution of α -methyl-glucopyranoside (5 g, 25.8 mmol) **1** in 50 mL pyridine at 0°C was added portionwise 4-Toluenesulfonyl chloride (6.4 g, 33.5 mmol). The mixture was stirred for 12 hours from 0°C to room temperature. Pyridine was removed under reduced pressure and the resulting oil was submitted to flash chromatography (CH₂Cl₂/MeOH gradient = 98/2; 95/5 ; 9/1) to give **2** as a white solid (4.66g, 56%).

¹H NMR (400 MHz, CDCl₃): δ (ppm) 7.82 (d, 2H, J = 8.4 Hz, 2H-8), 7.36 (d, 2H, J = 8.4 Hz, 2H-9), 4.60 (d, 1H, J = 3.7 Hz, H-1), 4.31 (dd, 2H, J = 4.6 Hz and J = 1.8 Hz, H-4, H-5), 3.73 (d, 2H, J = 9 Hz, H-3, H-6), 3.49 (s, 3H, OCH₃), 3.31 (d, 2H, J = 9.7 Hz, H-2, H-6), 2.46 (s, 3H, CH₃)

¹³C NMR (100 MHz, CDCl₃): δ (ppm) 143.9 (C7), 131.8 (C10), 128.9, 126.9 (C8, C9), 98.3 (C1), 73.3, 70.9, 68.9, 68.5 (4CH, C2-C5), 67.9 (C6), 54.4 (OCH₃), 20.6 (CH₃)

HRMS calcd for C₁₄H₂₁O₈S [M+H⁺]: 349.0951. Found: 349.0939

6-azido-6-deoxy- α -methyl-D-glucopyranoside (**3**)



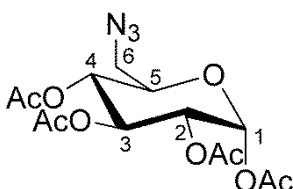
To a solution of **2** (4.66 g, 13.5 mmol) in 150 mL of water/acetone (2/3) was added sodium azide (8.78 g, 135 mmol). Reaction was stirred under reflux at 90°C for 12 hours. Solvents were removed under reduced pressure and dichloromethane was added to the resulting white residue. The solution was filtrated and the filtrate was concentrated under reduced pressure. The resulting solid was submitted to flash chromatography (CH₂Cl₂/MeOH gradient: 95/5 ; 9/1) to give **3** as a white solid (2.1 g, 75%).

¹H NMR (400 MHz, CDCl₃): δ (ppm) 4.63 (d, 1H, *J* = 8.7 Hz, H-1), 3.73 (d, 2H, *J* = 10.8 Hz, H-4, H-5), 3.58 (m, 2H, H-6a, H-3), 3.50 (s, 3H, OH) 3.42 (m, 2H, H-6b, H-2) 3.30 (s, 3H, OCH₃).

¹³C NMR (100 MHz, CDCl₃): δ (ppm) 99.4 (C1), 74.1, 71.9, 70.8, 70.7 (C2-C5), 55.4 (OCH₃), 51.3 (C6)

HRMS calcd for C₇H₁₄N₃O₅ [M+H⁺]: 220.0928. Found: 220.0922

1,2,3,4-tetra-*O*-acetyl-6-azido-6-deoxy-α-D-glucopyranoside (**4**)



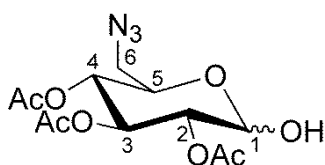
Compound **3** (1.83 g, 8.36 mmol) was dissolved in 170 mL of a Ac₂O:AcOH (1:1) mixture, and cooled to 0°C. 1.8 mL of sulfuric acid was then added dropwise. After 30 minutes of stirring, the temperature was set to 25 °C and the reaction was carried on for 8 more hours. To the solution was then added 200 mL of cold water (4°C). The solution was extracted three times with dichloromethane. Organic phases were combined, washed two times with water and two times with saturated NaHCO₃, and dried with MgSO₄. The solvent were removed under reduced pressure. The resulting product was submitted to flash chromatography (Cyclohexane/Ethyl acetate: 7/3) to give **4** as a white solid (2.37 g, 76 %).

¹H NMR (400 MHz, CDCl₃) : δ (ppm) 6.38 (d, 1H, *J* = 3.7 Hz, H-1), 5.51 (d, 1H, *J* = 10.08 Hz, H-3), 5.12 (m, 2H, H-2, H-4), 4.10 (m, 1H, H-5), 3.34 (m, 2H, H-6a, H-6b), 2.14 (s, 3H, CH₃COO(α)), 2.12 (s, 3H, CH₃COO), 2.09 (s, 3H, CH₃COO), 2.04 (s, 3H, CH₃COO).

¹³C NMR (100 MHz, CDCl₃): δ (ppm) 169.2, 168.6, 168.4, 167.7 (CH₃C=O), 87.7 (C1), 69.8, 68.7, 68.5, 68.1 (C2-C5), 49.6 (C6), 19.8, 19.7, 19.6, 19.4 (CH₃COO)

HRMS calcd for C₁₄H₁₉N₃NaO₉ [M+Na⁺]: 396.1013. Found: 396.1024

2,3,4-tri-*O*-acetyl-6-azido-6-deoxy-D-glucopyranose (**5**)



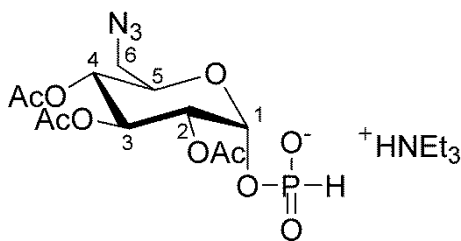
To a solution of **4** (1.23 g, 3.3 mmol) in 17 mL of anhydrous THF under nitrogen atmosphere was added benzylamine (432 μL; 3.96 μmol). The mixture was stirred at 50°C for 8 hours and overnight at 25°C. Solvents were then removed under reduced pressure and the resulting product was purified by flash chromatography (Petroleum ether/Ethyl acetate: 7/3) to give **5** as a yellow oil (984 mg, 90%).

¹H NMR (400 MHz, CDCl₃): δ (ppm) 5.52 (dd, 1H, *J* = 10.0 Hz, *J* = 3.1 Hz, H-3α), 5.46 (d, 0.60H, *J* = 2 Hz, H-1α), 5.27-5.15 (m, 2H, H-4α, H-2α), 5.12 (m, 0.4H, H-3β), 4.78 (d, 0.4H, *J* = 8.01 Hz, H-1β), 4.24 (m, 1H, H-5α), 3.34-3.28 (m, 2H, H-6α), 2.13 (s, 3H, CH₃COO), 2.02 (s, 3H, CH₃COO), 1.97 (s, 3H, CH₃COO)

¹³C NMR (100 MHz, CDCl₃): δ (ppm) 170.3, 169.8, 169.5 (3×CH₃C=O), 95.3 (C1α), 77.0 (C1β), 73.2 (C2α), 73.1 (C2β), 72.3 (C4α), 71.6 (C4β), 69.7 (C3β), 68.2 (C3α), 61.7 (C5β), 60.5 (C5α), 51.04 (C6α), 50.9 (C6β), 20.65, 20.64, 20.59 (3×CH₃COO)

HRMS calcd for C₁₂H₁₇N₃NaO₈ [M+Na⁺]: 354.0907. Found: 354.0944

2,3,4-tri-*O*-acetyl-6-azido-6-deoxy-α-D-glucopyranosyl phosphite (**6**)



To a solution of **5** (700 mg, 2.12 mmol) in 9 mL of an anhydrous THF/1.4 dioxane mixture was added triethylamine (456 μ L; 3.29 mmol). 2-chloro-4H-benzodioxaphosphorine-4-one (648 mg; 3.2 mmol) was solubilized in 2 mL of anhydrous THF. The two solutions were mixed under nitrogen atmosphere at 0°C and stirred for 30 minutes. Water (1.8 mL) was added and solvents were removed under reduced pressure. 40 mL of THF were then added and the brown precipitate was removed by filtration. The filtrate was concentrated under reduced pressure and purified by flash chromatography ($\text{CHCl}_3/\text{MeOH}/\text{NEt}_3$; 90/9.9/0.1) to give **6** as a yellow oil (384 mg, 46%).

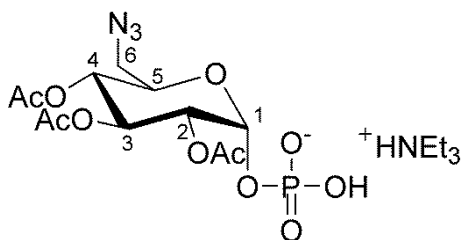
^1H NMR (400 MHz, CD_3OD) : δ (ppm) 6.79 (d, 1H, P-H), 5.78 (dd, 1H, $J_{\text{H-P}} = 8.8$ Hz, $J_{1-2} = 3.32$ Hz, H-1), 5.50 (t, 1H, $J = 9.7$ Hz, H-4), 5.34 (t, 1H, $J = 9.5$ Hz, H-3), 5.10 (t, 1H, $J = 9.9$ Hz, H-2), 4.27 (m, 1H, H-5), 3.52 (dd, 1H, $J = 7.12$ Hz, $J = 6.9$ Hz, H-6a), 3.34 (q, 6H, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 3.32 (dd, 1H, $J = 7.12$ Hz, $J = 6.96$ Hz, H-6b), 2.17 (s, 3H, CH_3COO), 2.05 (s, 3H, CH_3COO), 2.00 (s, 3H, CH_3COO), 1.19 (t, 9H, $J = 7.16$ Hz, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$)

^{13}C NMR (100 MHz, CD_3OD): δ (ppm) 170.34 (COOCH_3), 170.27 (COOCH_3), 169.91 (COOCH_3), 98.82 (C1), 70.66 (C5), 69.91 (C4), 68.77 (C2), 67.95 (C3), 54.65 (C6), 47.63 (3C, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 19.66 (CH_3COO), 19.53 (CH_3COO), 19.22 (CH_3COO), 9.18 (3C, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$)

^{31}P NMR (202.30 MHz, CD_3OD) : δ (ppm) -0.19 (s, 1P, P-H)

MS (ESI-negative mode) m/z 394.062 (M^-)

2,3,4-tri-*O*-acétyl-6-azido-6-désoxy- α -D-glucopyranosyl phosphate (**7**)



Compound **6** (380 mg; 964 μmol) was dissolved in 15 mL of THF and run through DOWEX (50W-X8 H^+ form). After filtration and removal of the solvent under reduced pressure, the resulting yellow oil was dissolved in 15 mL of anhydrous THF. *Tert*-butyl hydroxypoxide (214 μL ; 1.9 mmol) was then added and the mixture was stirred for 24 hours at 25 $^{\circ}\text{C}$. To neutralize the reaction, triethylamine was added (274 μL ; 1.9 mmol) and the solvent was removed under reduced pressure. The resulting oil was purified by reversed phase chromatography (C18 phase) with water as the eluent. Isolated fractions were lyophilized to give **7** as a yellow oil (292 mg, 74%).

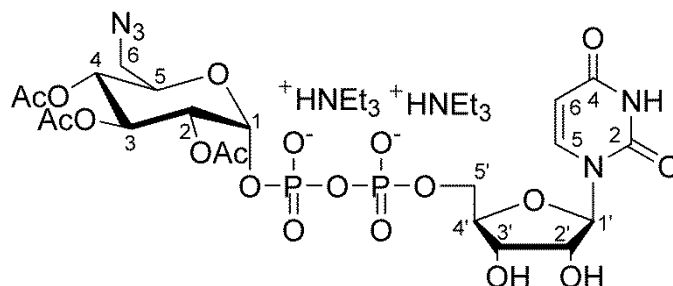
^1H NMR (400 MHz, CD_3OD): δ (ppm) 5.75 (dd, 1H, $J_{\text{H-P}} = 8.8$ Hz, $J_{1-2} = 5.48$ Hz, H-1), 5.53 (t, 1H, $J = 10.0$ Hz, H-4), 5.50 (t, 1H, $J = 9.7$ Hz, H-3), 4.89 (m, 1H, $J = 9.9$ Hz, H-2), 4.31 (m, 1H, H-5), 3.59 (dd, 1H, $J = 10.5$ Hz, $J = 3.1$ Hz, H-6a), 3.32 (dd, 1H, $J = 10.5$ Hz, $J = 3.1$ Hz, H-6b), 3.30 (q, 6H, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 2.06 (s, 3H, CH_3COO), 2.04 (s, 3H, CH_3COO), 2.00 (s, 3H, CH_3COO), 1.30 (t, 9H, $J = 7.1$ Hz, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$)

^{13}C NMR (100 MHz, CD_3OD): δ (ppm) 170.41 (COOCH_3), 170.29 (COOCH_3), 169.86 (COOCH_3), 91.5 (C1), 70.63 (C5), 70.04 (C4), 69.13 (C2), 68.82 (C3), 50.29 (C6), 48.95 (3C, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$), 19.24 (CH_3COO), 19.19 (CH_3COO), 19.16 (CH_3COO), 7.77 (3C, $(\text{CH}_3\text{CH}_2)_3\text{NH}^+$)

^{31}P NMR (202.30 MHz, CD_3OD): δ (ppm); 1.63 (s, 1P, P-OH)

MS (ESI-negative mode) m/z 410.0 (M^-)

Uridine 5'-(2,3,4-tri-*O*-acetyl-6-azido-6-deoxy- α -D-glucopyranosyl) diphosphate (**8**)



Phosphate **7** (149 mg, 360 μmol) was co-evaporated with dry pyridine (10 mL) for three times under reduced pressure. UMP-morpholidate (495 mg; 720 μmol) and tetrazole (1.6 mL; 720 μmol , 0.45 M in acetonitrile) were co-evaporated in another flask with dry pyridine (5 mL) for three times under reduced pressure. Compound **7** was then added to the mixture with 10

mL pyridine, and the mixture was co-evaporated with dry pyridine (10 mL) for two times under reduced pressure. Dry pyridine was then added (20 mL), partially evaporated and the mixture was stirred at rt under argon atmosphere in the dark. After two days, toluene (40 mL) was added and solvents were removed under reduced pressure. The residue was then lyophilized and purified by C18 reverse-phase HPLC, eluting with 0-60% CH₃CN in 0.04 M TEAA in 60 minutes(4mL/min). After lyophilization, compound **8** was obtained as a pale yellow oil (120 mg, 36%).

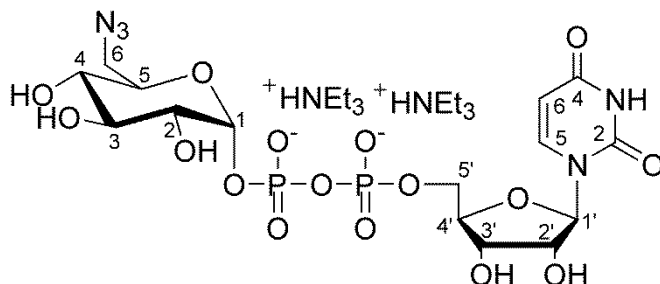
¹H NMR (400 MHz, D₂O): δ (ppm) 7.96 (d, 1H, J = 8.08 Hz, H₆), 5.92 (d, 1H, J = 8.08 Hz, H₅), 5.89 (d, 1H, J 1-2 = 3,4 Hz, H_{1'}), 5.67 (dd, 1H, J 1-2 = 3,28 Hz, H₁-glu), 5.36 (m, 1H, J = Hz H₄-glu), 5.10 (m, 1H, J = Hz, H₃-glu), 4.98 (m, 1H, J =9.88 Hz, H₂-glu), 4.27 (m, 1H, H₅-glu), 4.23 (m, 2H, H_{2'}-H_{3'}), 4.19 (m, 2H, H_{4'}-H_{5'a}), 4.11 (m, 1H, H_{5'b}), 3.53 (dd, 1H, H_{6a}-glu), 3.46 (dd, 1H, H_{6b}-glu), 3.11 (q, 12H, (CH₃CH₂)₃NH⁺), 2.04 (s, 3H, (CH₃COO)), 1.99 (s, 3H, (CH₃COO)), 1.96 (s, 3H, (CH₃COO)), 1.18 (t, 18H, (CH₃CH₂)₃NH⁺)

¹³C NMR (100 MHz, D₂O): δ (ppm) 170.41(COOCH₃), 170.29 (C=O), 169.86 (C=O), 164.2 (C₄), 150.4 (C₂), 142.3 (C₆), 104.9 (C₅), 91.5 (C₁), 88.1 (C_{1'}), 84.2 (C_{4'}), 73.7 (C_{2'}), 70.8 (C_{3'}), 70.63 (C₅-glu), 70.04 (C₄-glu), 69.13 (C₂-glu), 68.82 (C₃-glu), 65.6 (C_{5'}), 50.29 (C₆-glu), 48.95 (6C, (CH₃CH₂)₃NH⁺), 19.24 (CH₃COO), 19.19 (CH₃COO), 19.16 (CH₃COO), 7.77 (6C, (CH₃CH₂)₃NH⁺).

³¹P NMR (202.30 MHz, D₂O): δ (ppm) -8.9 (d, 1P, J_{p-p} = 19.6 Hz, P-rib), -11.2 (d, 1P, J_{p-p} = 19.6 Hz, P-glu)

MS (ESI-negative mode) m/z 716.1 (M)⁻

Uridine 5'-(6-azido-6-deoxy- α -D-glucopyranosyl) diphosphate (**9**)



The compound **8** (111 mg, 120 μ mol) was dissolved in methanol (8.4 mL) at 0°C, then an aqueous solution of NH₄HCO₃ (0.1 M, 11.1 mL) and triethylamine (432 μ L, 3120 μ mol) were added and the mixture was stirred at 0°C for 23 hr. Then water (37.5 mL) was added and pH

was adjusted to 7.5 with DOWEX 50W (H⁺ form) resin. The resin was removed by filtration through a PTFE filter and washed with water (20 mL). The filtrate was lyophilized and purified by C18 reverse-phase HPLC, eluting with 0-25% CH₃CN in 0.04 M TEAA in 30 minutes (4mL/min). After lyophilization, compound **9** was obtained as a white powder (80 mg, 85%).

¹H NMR (500 MHz, D₂O): δ (ppm) 7.96 (d, 1H, J = 8.1 Hz, H6-ura), 6.00-5.96 (m, 2H, H1', H5-ura), 5.58 (dd, 1H, J = 7.1, 3.4 Hz, H1-glu), 4.37-4.35 (m, 2H, H3', H2'), 4.28-4.18 (m, 3H, H5'a, H5'b, H4'), 4.02 (dd, 1H, J = 6.7, 3.1 Hz, H4-glu), 3.78-3.70 (m, 2H, H6-glu a, H5glu), 3.61-3.47 (m, 3H, H6-glu b, H3-glu, H2-glu), 3.19 (q, 12H, J = 7.3 Hz, (CH₃CH₂)₃NH⁺), 1.27 (t, 18H, J = 7.3 Hz, (CH₃CH₂)₃NH⁺).

¹³C NMR (125 MHz, D₂O): δ (ppm) 168.9 (C4-ura), 154.5 (C2-ura), 144.0 (C6-ura), 104.9 (C5-ura), 97.7 (C1-glu), 90.7 (C1'), 85.5 (C4'), 76.0 (C2'), 74.9 (C5-glu), 73.8 (C3-glu), 73.7 (C4-glu), 71.8 (C3'), 72.0 (C2-glu), 67.1 (C5'), 52.7 (C6-glu), 48.9 (CH₃CH₂)₃NH⁺), 10.5 (CH₃CH₂)₃NH⁺).

³¹P NMR (202.30 MHz, D₂O): δ (ppm) -8.7 (d, 1P, J_{p-p} = 22.1 Hz, P-rib), -10.5 (d, 1P, J_{p-p} = 19.4 Hz, P-glu)

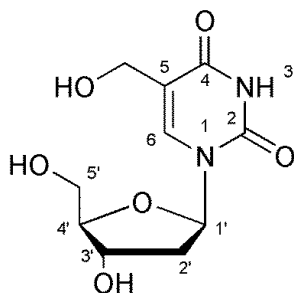
HRMS for C₁₅H₂₁N₅O₁₆P₂²⁻ [M]: calcd 589.0469, found 589.0441

2. Synthèse de la 5-hydroxyméthyl-2'-désoxycytidine

Les analyses par HPLC ont été effectuées sur une HPLC Waters avec une colonne XTerra C18 (5 μ M, 4,6 x 250 mm). La phase mobile est un mélange d'eau (avec 0.1% de TFA) et de méthanol (avec 0.1% de TFA). Un volume de 20 μ L est injecté. Un gradient linéaire de 40% de méthanol est appliqué en 30 min avec un débit de 1mL/min. L'absorbance est suivie à 260 nm.

Les purifications par HPLC ont été effectuées sur une HPLC Waters avec une colonne Xterra Prep RP18 (10 μ M, 7,8 x 300 mm). La phase mobile est un mélange d'eau (avec 0.1% de TFA) et de méthanol (avec 0.1% de TFA). Un volume de 20 μ L est injecté. Un gradient linéaire de 15 à 50% de méthanol est appliqué en 45 min avec un débit de 3mL/min. Les fractions recueillies sont ensuite analysées par spectrométrie de masse.

5-Hydroxyméthyl-2'-désoxyuridine (11)



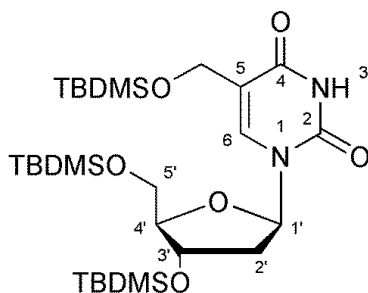
La 2'-désoxyuridine (0.9 g, 3.95 mmol) est solubilisée dans l'eau (10 mL), puis est ajouté de la triéthylamine (7 mL, 51 mmol) et du formaldéhyde 37% (18 mL, 222 mmol). Le mélange est porté à reflux pendant 3 jours à 80°C puis pendant 1 jour à 100°C. L'eau est ensuite évaporée et le produit obtenu est soumis à une chromatographie sur gel de silice (2-15% méthanol dans dichlorométhane). Le produit **11** est obtenu, sous la forme d'une poudre blanche, avec un rendement de 60% (0.625 g, 2.4 mmol).

¹H NMR (DMSO-d₆, 400 MHz) δ ppm: 2.06 (m, 2H, H-2' et H-2''); 3.54 (m, 2H, H-5' et H-5''); 3.77 (m, 1H, H-4'); 4.13 (m, 2H, CH₂-OH); 4.24 (m, 1H, H-3'); 4.89-4.96 (2m, 2H, OH-5' et CH₂-OH-5); 5.24 (m, 1H, OH-3'); 6.19 (t, 1H, *J* = 6.6 Hz, *J* = 7.2 Hz, H-1'); 7.74 (s, 1H, H-6).

¹³C NMR (DMSO-d₆, 125 MHz) δ ppm: 40.18 (C-2'); 61.24 (C-5-CH₂); 61.60 (C-5'); 70.69 (C-3'); 84.06 (C-1'); 87.45 (C-4'); 114.42 (C-5); 136.91 (C-6); 150.52 (C-2); 162.79 (C-4)

HRMS (EI) [M+Na⁺] *m/z* Calculé: 281.07469. Trouvé: 281.07441

3',5'-Di-*O*-*tert*-butyldiméthylsilyl-5-*O*-(*tert*-butyldiméthylsilyl)-2'-d désoxyuridine (12)



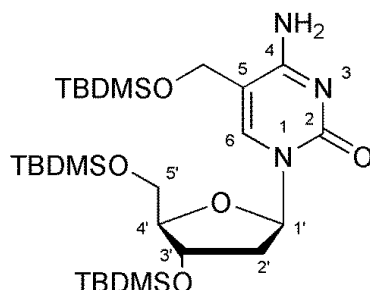
Le composé **11** (0.3 g, 1.1 mmol) est agité à température ambiante pendant 3 jours avec de l'imidazole (0.33 g, 4.7 mmol), du TBDMSCl (0.73 g, 4.7 mmol) dans du DMF (2.5 mL). De

l'eau (20 mL) est ajoutée et la phase aqueuse est extraite avec de l'acétate d'éthyle (2x20 mL). Les phases organiques sont rassemblées et séchées avec du MgSO₄. Le produit obtenu est soumis à une chromatographie sur gel de silice (1-15% méthanol dans dichlorométhane). La 3',5'-Di-*O-tert*-butyldiméthylsilyl-5-*O*-(*tert*-butyldiméthylsilyl)-2'-désoxyuridine **12** est obtenue, sous la forme d'une huile jaune, avec un rendement de 36% (0.2423 g, 0.4 mmol).

¹H NMR (CDCl₃, 400 MHz) δ ppm: 0.07-0.12 (m, 18H, 6xCH₃-Si) ; 0.91-0.94 (m, 27H, 3 tBu) ; 2.03 (m, 1H, H-2'') ; 2.32 (m, 1H, H-2') ; 3.68 (m, 1H, H-5'') ; 3.87 (m, 1H, H-5') ; 3.96 (m, 1H, H-4') ; 4.38 (m, 1H, H-3') ; 4.47 (m, 2H, CH₂-O-5) ; 6.29 (dd, 1H, H-1', $J=5.5$ Hz, $J=7.9$ Hz) ; 7.52 (m, 1H, H-6) ; 8.29 (m, 1H, NH-3)

HRMS (EI) [M+Na⁺] m/z Calculé: 623.33408. Trouvé: 623.33384

3',5'-Di-*O-tert*-butyldiméthylsilyl-5-*O*-(*tert*-butyldiméthylsilyl)-2'-désoxycytidine (**13**)

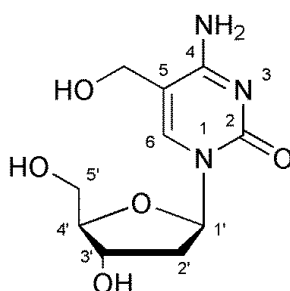


Le composé **12** (0.288 g, 0.47 mmol) est solubilisée dans de l'acétonitrile (10 mL). La DMAP (0.173 g, 1.4 mmol), la triéthylamine (2.5 mL, 16 mmol) et le chlorure de triisopropylbenzenesulfonyle (0.429 g, 1.4 mmol) sont ajoutés et le mélange réactionnel est agité à température ambiante pendant 21h. De l'ammoniaque à 28% (12 mL) est ensuite ajouté à la réaction qui se poursuit pendant 4h à température ambiante. L'acétonitrile est évaporé puis de l'eau (40 mL) est ajoutée et la phase aqueuse est extraite avec de l'acétate d'éthyle (2x40 mL). Les phases organiques sont rassemblées et séchée avec du MgSO₄. Le produit obtenu est soumis à une chromatographie sur gel de silice (0-2% méthanol dans du dichlorométhane). La 3',5'-Di-*O-tert*-butyldiméthylsilyl-5-*O*-(*tert*-butyldiméthylsilyl) -2'-désoxycytidine **13** est obtenue, sous la forme d'une poudre brune, avec un rendement de 67 % (0.191 g, 0.32 mmol).

¹H NMR (CDCl₃, 400 MHz) δ ppm: 0.04-0.08 (m, 18H, 6xCH₃-Si) ; 0.86-0.90 (m, 27H, 3 tBu) ; 1.92 ((m, 1H, H-2'')) ; 2.42 (ddd, 1H, H-2', *J* = 3.4 Hz, *J* = 6.1 Hz, *J* = 13.5 Hz) ; 3.74 (dd, 1H, H-5'', *J* = 3.1 Hz, *J* = 11.2 Hz) ; 3.8 (dd, 1H, H-5', *J* = 3.3 Hz, *J* = 11.3 Hz) ; 3.92 (q, 1H, H-4', *J* = 3.1 Hz) ; 4.32 (m, 1H, H-3') ; 4.43 (s, 2H, CH₂-O-5) ; 6.28 (t, 1H, H-1', *J* = 6.3 Hz, *J* = 6.6 Hz) ; 7.59 (s, 1H, H-6).

HRMS (EI) [M+Na⁺] *m/z* Calculé: 622.34995. Trouvé: 622.34982

5 -Hydroxyméthyl-2'-désoxycytidine (5hmdC)



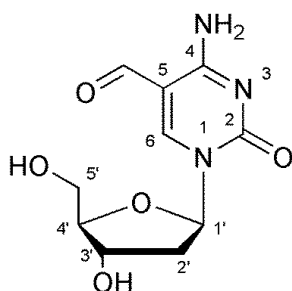
Le composé **13** (0.078 g, 0.13 mmol) est solubilisé dans du THF (1.5 mL) auquel est ajouté du TBAF.3H₂O (0.123 g, 0.39 mmol) pendant environ 2h à température ambiante. Le mélange réactionnel est ensuite évaporé puis purifié sur colonne de silice (MeOH dans CH₂Cl₂ : 5%-20%). La 5-hydroxyméthyl-2'-désoxycytidine est obtenue, sous la forme d'une poudre beige, avec un rendement de 77% (0.026 g, 0.10 mmol).

¹H NMR (DMSO-d₆, 400 MHz) δ ppm: 1.94 (m, 1H, H-2'') ; 2.09 (m, 1H, H-2') ; 3.55 (m, 2H, H-5' et H-5'') ; 3.76 (q, 1H, H-4', *J* = 3.9 Hz) ; 4.18 (m, 2H, CH₂-OH-5 et H-3') ; 4.97 (m, 3H, CH₂-OH-5 et OH-5') ; 5.18 (d, 1H, OH-3', *J* = 4.3 Hz) ; 6.17 (t, 1H, H-1', *J* = 6.9 Hz) ; 6.57 (brs, 1H, NH) ; 7.3 (brs, 1H, NH) ; 7.74 (s, 1H, H-6).

¹³C NMR (DMSO-d₆, 125 MHz) δ ppm: 40.28 (C-2') ; 57.46 (C-5-CH₂) ; 61.38 (C-5') ; 70.40 (C-3') ; 84.69 (C-1') ; 87.12 (C-4') ; 105.90 (C-5) ; 138.91 (C-6) ; 154.94 (C-2) ; 164.38 (C-4)

HRMS (EI) [M+Na⁺] *m/z* Calculé: 280.09060. Trouvé : 280.09039

5 -Formyl-2'-désoxycytidine (5fdC)

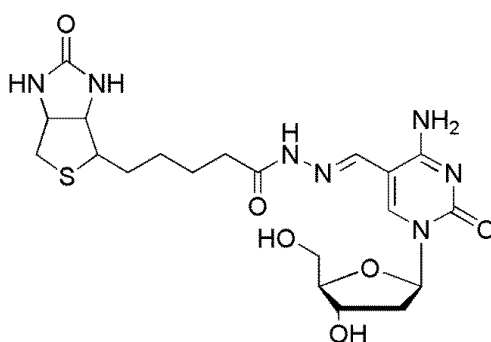


La 5-hydroxyméthyl-2'-désoxycytidine (0.008 g, 0.031 mmol) est solubilisée dans de l'eau (0.5 mL) et du dioxane (0.5 mL) puis est ajouté de l'oxyde de ruthénium (0.021 g, 0.16 mmol). Le mélange réactionnel est mis sous agitation à 60°C pendant 4h puis est filtré sur des Ministart 0,45µM (Sartorius Stedim biotech) et lyophilisé. Le brut réactionnel est analysé puis purifié par HPLC dans les conditions décrites page 238. Le temps de rétention de ce composé oxydé dans ces conditions est de 17 min, celui de la 5hmdC est de 5 min.

¹H NMR (DMSO-d₆, 400 MHz) δ ppm: 2.09 (m, 1H, H-2'') ; 2.30 (m, 1H, H-2') ; 3.63 (2m, 2H, H-5'' et H-5') ; 4.17 (s, 1H, CH₂-OH) ; 4.22 (s, 1H, H-3') ; 5.14-5.28 (2m, 2H, OH-3' et OH-5') ; 6.09 (t, 1H, H-1', ³J_{H1'-H2'} = 5.9 Hz) ; 7.86-8.17 (2s, 2H, NH₂) ; 8.84 (s, 1H, H-6) ; 9.46 (s, 1H, CHO).

MS (EI) *m/z*: 256.1 (M+H⁺); 140.1 (5fC + H⁺)

5 - Biotine hydrazone-2'-désoxycytidine (5biotdC)



La 5-Formyl-2'-désoxycytidine (3 mg, 0.012 mmol) est solubilisée dans l'eau en présence de 10% d'acide acétique. Après ajout de l'hydrazine biotinylée (6 mg, 0.024 mmol). Le mélange réactionnel est ensuite placé à l'abri de la lumière à 37°C sous agitation durant 12h. L'acide acétique est ensuite évaporé et le mélange lyophilisé. Le brut réactionnel est analysé puis purifié par HPLC dans les conditions décrites page 238.

MS (EI) *m/z* : 496.4 (M+H⁺) ; 380.4 (5biotC + H⁺)

3. Essais de fonctionnalisation chimique sur l'oligonucléotide H

Traitement des échantillons

L'oligonucléotide H (**Tableau 8**) (1 μL , 9.9 nmol) est solubilisé dans 19 μL d'eau en présence de RuO_2 (1 mg, 7,5 μmol) pendant 4h à 60°C. Le mélange réactionnel est ensuite solubilisé dans 22 μL d'eau et mis en présence d'hydrazine biotinylée (10 μL , 50 μM). La solution est complétée à 50 μL avec de l'acide acétique, afin d'avoir une solution à 10% en acide. Le mélange est incubé à 37°C dans l'obscurité pendant 12h puis lyophilisé pendant 12h.

Purification des oligonucléotides sur colonne d'exclusion stérique

Les colonnes Bio-Spins 6 ou 30 ont été utilisées selon les recommandations du fournisseur. Brièvement, elles sont lavées 3 fois à l'eau afin d'éliminer le tampon Tris de stockage et subissent plusieurs cycles de centrifugation (2 min à 1.000g). Puis, 20 μL d'oligonucléotide oxydé par du RuO_2 sont déposés sur les colonnes. Après 4 min de centrifugation à 1.000g, l'éluat contenant oligonucléotide est récolté et sa concentration est évaluée sur un spectrophotomètre NanoDrop à 260 nm.

Purification des oligonucléotides par affinité sur bille de streptavidine

La recapture des oligonucléotides après biotinylation a été réalisée à l'aide de billes magnétiques greffées à la streptavidine (Dynabeads MyOne Streptavidin C1, 10 mg/ml), selon les recommandations du fournisseur. L'oligonucléotide (5 μM) est incubé avec 20 μL de billes prélavées par un tampon Tris-HCl (5 mM), EDTA (0.5 mM) et NaCl (1 M), pendant 15 minutes sous agitation. Les billes sont ensuite aimantées et lavées deux fois avec le même tampon 2 fois plus concentré. Puis l'oligonucléotide est relargué grâce à une solution aqueuse à 0.1% en TFA.

Digestion enzymatique de l'oligonucléotide purifié

16 μL de l'oligonucléotide (2 μM) purifié ont été traités. L'échantillon est mis à 90°C pendant 3 min puis placé dans la glace, 2 μL de tampon AcONH_4 (0,1 M, pH 5,3) puis 1 μL de nucléase P1 (2 U/ μL) sont ajoutés. L'échantillon est ensuite incubé à 45°C pendant 2h. 2.2 μL de tampon NH_4HCO_3 à 1 M ainsi que 1 μL de phosphodiesterase (0,002 U/ μL) sont ajoutés et

l'échantillon est de nouveau incubé à 37°C pendant 2h puis 0.5 µL de phosphatase (5 U/µL) sont ajoutés. L'échantillon est incubé à 37°C pendant 1h.

Analyse par Spectrométrie de masse

Les analyses par spectrométrie de masse ont été effectuées avec un couplage LC-MS constitué d'un système HPLC Dionex Ultimate 3000 (micropompe LG3400 + autosampler WPS3000) et d'un spectromètre de masse tandem Applied Biosystems QTRAP à travers une interface de source d'ions Turbolonspray.

10 µL de chaque échantillon (2 µM) ont été injecté dans une colonne, en phase inverse, de type Phenomenex Synergi 4u Polar et élué avec un tampon A (0.1% d'acide formique dans l'eau) et B (0.1% d'acide formique dans du méthanol) avec un gradient augmentant de 2%/min de 0 à 40% de B.

Les conditions de l'ESI sont les suivantes : flux de gaz de nébuliseur GS1 = 20, flux de gaz de rideau CUR = 20, flux de gaz de collision CAD = high, tension IonSpray IS = +5500V, potentiel Declustering DP = 20V, potentiel d'entrée EP = 10V, énergie de collision CE = 10V, collision cell exit potential CXP=15 V. Les chromatogrammes UV ont été effectués à $\lambda = 260$ nm. Les chromatogrammes LC-ESI-MS/MS ont été acquis en mode MRM (Multiple Reaction Monitoring) en contrôlant 7 paires d'ion moléculaire/ion fragment à m/z 228.1/112.1, 243.1/127.1, 252.2/136.2, 256/140, 258.1/142.1, 268.1/152.1, 496.4/380.4 correspondant respectivement à dC, dT, dA, 5fC, 5hmdC, dG et 5biotC. Le temps de scan a été mis à 200 ms pour chaque paire d'ions.

4. Essais de fonctionnalisation enzymatique

Obtention de la BGT

La BGT (sauvage et mutante E22A) a été purifiée sur une colonne SP Sepharose (Tampon A 10 mM KH_2PO_4 pH=7.5, 50 mM NaCl, 2 mM EDTA ; Tampon B 10 mM KH_2PO_4 pH=7.5, 1 M NaCl, 2 mM EDTA ; 4mL/min ; 25%B pendant 40 minutes puis 100%B pendant 20 minutes) puis sur une colonne Hiload 26/60 Superdex 200 (Tampon Tris 50 mM pH=7.8, 20 mM MgCl_2 , 4mL/min), comme décrit précédemment¹⁷⁷.

Formation du duplex.

Le duplex Hds-Ctrl a été formé en mélangeant des quantités équimolaires du brin H et du brin Ctrl (Tableau X) dans une solution de 150 mM NaCl et 10 mM Tris pH=7.4, puis en chauffant cette solution 5 minutes à 95°C et la laissant refroidir pendant la nuit.

Essais enzymatiques

L'enzyme BGT (0.15 μM final) est ajoutée dans une solution contenant 150 μM de l'UDP-6-azido-glucose et 30 μM du duplex Hds-Ctrl dans un tampon 50 mM HEPES pH=7.9, 25 mM MgCl₂. Après incubation pendant 2 heures à 37°C, le duplex a été isolé à l'aide d'une petite colonne d'exclusion stérique BioSpin P6 (Biorad). Les échantillons ont ensuite été soit directement analysés par MALDI-TOF ou par la méthode de LC-MS/MS décrite précédemment, soit séparés par HPLC (C18, 1 mL/min, 0 à 40 % d'acétonitrile dans 0.1 M TEAA en 60 minutes détection à 260nm) et analysés ensuite par MALDI-TOF.

Réaction de chimie « click »

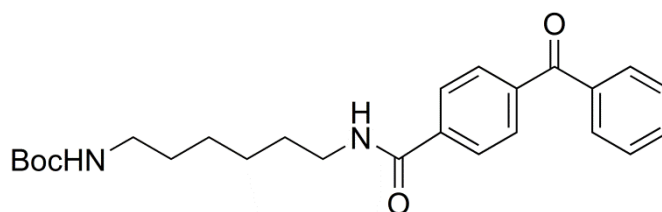
Le composé 3 (2 μM final) et le composé DBCO-Alexa Fluor 488 (2 μM final) sont dissous dans 50 μL d'eau. La réaction est incubée à 37°C pendant deux heures puis est analysée par HPLC (C18, 1 mL/min, 0 à 60 % d'acétonitrile dans 0.1 M TEAA en 40 minutes, détection à 488 nm). Les pics d'intérêts ont été recueillis et lyophilisés puis analysés par MS/MS.

Partie expérimentale du chapitre IV

Oligodeoxynucleotide synthesis was done at the Institut de Chimie de Nice (Nice, France). MALDI-TOF spectra of oligonucleotides were recorded at the Institut de Chimie des Substances Naturelles (Gif sur Yvette, France). All experiments involving radioactivity were done at the Laboratoire Structure et Instabilité du Génome, Muséum d'Histoire Naturelle (Paris, France). Amine synthesis, oligodeoxynucleotide fonctionnalisation and purification, and measurement of the melting temperature were done at the Laboratoire des Biomolécules (Paris, France).

1) Synthesis of photoactivable amines

[(4-Benzoyl-benzoylamino)hexyl]carbamic acid *tert*-butyl ester (**57**)



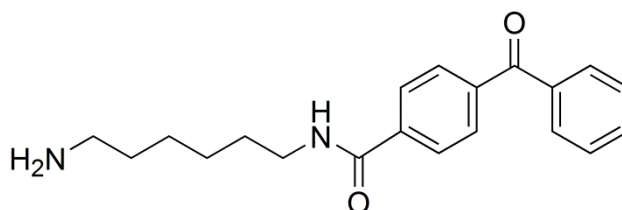
To a stirred solution of 4-benzoylbenzoic acid (2.2 mmol, 500 mg) in DMF (10 mL) were added DIPEA (1.1 equiv), and HATU (1.1 equiv). The mixture was stirred for 20 min and a solution of *tert*-butyl-*N*-(6-aminohexyl)carbamate (2.2 mol, 556 mg) previously dissolved in DMF (5 mL) in the presence of DIPEA (1.2 equiv) was added to the ester. The resulting mixture was stirred at room temperature in the dark overnight. The solvents were removed under reduced pressure and the crude residue was purified by flash chromatography (CH₂Cl₂/MeOH) to give **57** as a yellow solid (825 mg, 88%).

¹H-NMR (400 MHz, CDCl₃): δ (ppm) 7.91 (dd, 2H); 7.82 (dd, 2H); 7.78 (dd, 2H); 7.60 (m, 2H); 7.48 (m, 1H); 6.66 (br, 1H); 3.47 (dd, 2H); 3.12 (dd, 2H); 1.64 (m, 2H); 1.49 (m, 2H); 1.43 (m, 2H); 1.42 (s, 9H); 1.37 (m, 2H)

¹³C-NMR (100 MHz, CDCl₃): δ (ppm) 196.1; 166.7; 156.4; 140.0; 138.1; 137.1; 132.9; 130.1; 128.5; 127.0; 40.1; 39.7; 30.1; 29.4; 28.5; 26.0; 25.8.

HRMS calcd for C₂₅H₃₂N₂O₄Na [MNa⁺]: 447.2254 Found: 447.2257

***N*-Aminoethyl-4-benzoyl-benzamide (58)**



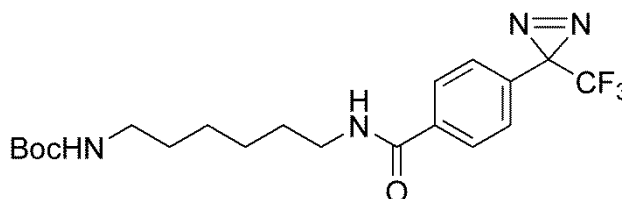
To a solution of **57** (800 mg, 1.88 mmol) in 10 mL of chloroform was added 4 mL of trifluoroacetic acid. The mixture was stirred at room temperature overnight in the dark. The solution was concentrated under reduced pressure and the crude residue was partitioned between 1 M NaOH and ethyl acetate. The aqueous layer was extracted 3 times with ethyl acetate. The combined organic layers were dried over Na₂SO₄ and then concentrated. The residue was purified by flash chromatography (CHCl₃/MeOH) to give **58** as a white solid (580 mg, 95 %).

¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) 8.70 (t, 1H); 7.99 (dd, 2H); 7.83 (br, 1H); 7.79 (dd, 2H); 7.75 (dd, 2H); 7.70 (dd, 1H); 7.58 (dd, 2H); 3.28 (m, 2H); 2.78 (m, 2H); 1.55 (m, 4H); 1.34 (m, 4H)

¹³C-NMR (100 MHz, DMSO-d₆): δ (ppm) 195.4; 165.4; 139.0; 138.0; 136.7; 133.0; 129.7; 129.5; 128.7; 127.3; 39.2; 38.7; 28.9; 27.0; 26.0; 25.5.

HRMS calcd for C₂₀H₂₅N₂O₂ [MH⁺]: 325.1910 Found: 325.1911

***N*-[3-[[4-[3-(trifluoromethyl)-3*H*-diazirin-3-yl]benzoyl]amino]hexylcarbamic acid *tert*-butyl ester (60).**



To a stirred solution of 4-[3-trifluoromethyl)-3*H*-diazirine-3-yl]benzoic acid (2.17 mmol, 500 mg) in DMF (12 mL) were added DIPEA (1.2 equiv), and HATU (1.2 equiv). The mixture

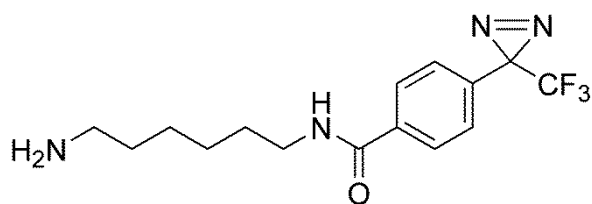
was stirred for 20 min and a solution of *tert*-butyl-*N*-(6-aminohexyl)carbamate previously dissolved in DMF (3 mL) in the presence of DIPEA (1.2 equiv) was added to the ester. The resulting mixture was stirred at room temperature in the dark overnight. The solvents were removed under reduced pressure and the crude residue was purified by flash chromatography (CH₂Cl₂/MeOH) to give **60** as a yellow solid (740 mg, 80%).

¹H-NMR (400 MHz, CDCl₃): δ (ppm) 7.85 (dd, *J* = 8.60 Hz, 2H); 7.23 (d, *J* = 8.25 Hz, 2H); 6.67 (br, 1H); 3.43 (m, 2H); 3.11 (t, 2H); 1.61 (m, 2H); 1.48 (m, 2H); 1.42 (s, 9H); 1.36 (m, 4H)

¹³C-NMR (100 MHz, CDCl₃): δ (ppm) 166.3; 156.3; 135.9; 132.1; 127.5; 126.6; 123.0; 79.3; 40.0; 39.6; 30.1; 29.3; 28.5; 28.4; 25.9; 25.7.

HRMS calcd for C₂₀H₂₇F₃N₄NaO₃ [MNa⁺]: 451.1927 Found: 451.1931

***N*-(3-aminohexyl)-4-[3-(trifluoromethyl)-3*H*-diazirin-3-yl]benzamide (61)**



To a solution of **3** (700 mg, 1.6 mmol) in 10 mL of chloroform was added 4 mL of trifluoroacetic acid. The mixture was stirred at room temperature overnight in the dark. The solution was concentrated under reduced pressure and the crude residue was partitioned between 1 M NaOH and ethyl acetate. The aqueous layer was extracted 3 times with ethyl acetate. The combined organic layers were dried over Na₂SO₄ and then concentrated. The residue was purified by flash chromatography (CHCl₃/MeOH) to give **61** as an oil (quantitative).

¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) 8.64 (t, *J* = 5.6 Hz, 1H) ; 7.95 (d, *J* = 8.5 Hz, 2H); 7.76 (br, 2H); 7.37 (d, *J* = 8.1 Hz, 2H); 3.26 (m, 2H); 2.76 (m, 2H); 1.51 (m, 4H); 1.31 (m, 4H)

¹³C-NMR (100 MHz, DMSO-d₆): δ (ppm) 165.0; 136.2; 130.0; 128.1; 126.4; 121.8; 39.1; 38.7; 28.8; 28.04; 27.0; 25.9; 25.5

HRMS calcd for C₁₅H₂₀F₃N₄O [MH⁺]: 329.1583 Found: 329.1585

2) Oligodeoxynucleotide synthesis and characterisation

Reagents and solvents for oligonucleotide synthesis, as well as dT, Ac-dC, Pac-dA, and dmF-dG phosphoramidites were purchased from Link Technologies. 2-F-dI, 5-I-dU and 3'-Protected BiotinLC Serinol CPG were purchased from GlenResearch. Unmodified ODNs and 4-thio-dT containing ODNs were purchased from Eurogentec (Belgium).

Oligonucleotides HPLC purifications were performed on a Waters HPLC apparatus (Waters™ 1525 Binary HPLC Pump with Waters™ 2487 Dual λ Absorbance Detector) using a Xterra Prep RP18 column (7.8*300 mm, 10 μ m) and purity was checked using a Waters Xterra MS C18 column (4.6*250 mm, 5 μ m). Absorbance spectrophotometry was performed on a Nanodrop-2000c spectrophotometer (Thermo Scientist).

Oligodeoxyribonucleotides synthesis

ODN synthesis was performed on an Expedite 8900 DNA synthesizer with Multiple oligo synthesis system (both from Applied Biosystem) using the “trityl on” mode and mild phosphoramidite chemistry on a 0.2 μ mol scale. The standard DNA assembly protocol “DMTr-on” was used except for the following modifications: 5-Ethylthio-1H-tetrazole (ETT) was used as activating agent; Pac-anhydride was used for capping; a longer coupling time was applied to 2-F-dI and 5-I-dU phosphoramidites (1200s) and to the following base (600s). Columns were flushed with argon after completion of the synthesis and stored at -20°C before post-synthetic modification.

Post-synthetic modification of 2-F-dI

Columns were treated with a 0.2-0.5M solution of the desired amine in and 1 equivalents of DIEA for 48-72 hours at 55°C. Columns were then deprotected with DBU and washed according to the instructions of the manufacturer. After that, ODN were cleaved from the solid support and deprotected overnight with concentrated aqueous ammonia at room temperature.

Oligodeoxyribonucleotides purification

The following gradient system was used: 95% A \rightarrow 60% A / 40% B \rightarrow 100% B \rightarrow 100%A with A = 0.1M TEAA pH 7.0 and B = Acetonitrile. After HPLC purification, ODNs were detritylated in 80% acetic acid for 1 hour at 22° C. The mixture was evaporated, 1,5 ml of water was added and the aqueous phase was extracted 3 times with Et₂O (3 \times 1.5 mL). ODNs were obtained after further HPLC purification and lyophilisation.

MALDI – TOF characterization of ODNs

A PerSeptive Voyager DE-STR MALDI-TOF mass spectrometer (PerSeptive Biosystems, Framingham, MA, USA), equipped with a 337-nm pulsed nitrogen laser (20 Hz) and an Acqiris® 2 GHz digitizer board, was used for all experiments. Mass spectra were obtained in the linear negative ion mode with the following settings: accelerating voltage 20kV, grid voltage 91 % of accelerating voltage, extraction delay time of 300 ns. The laser intensity was set just above the ion generation threshold to obtain peaks with the highest possible signal-to-noise (S/N) ratio without significant peak broadening. All data were processed using the Data Explorer software package (Applied Biosystems). 3-Hydroxypicolinic acid (3-HPA, used as the matrix for MALDI-TOF experiments, was of the highest grade available and used without further purification) was purchased from Sigma Aldrich Co.

Sequence	Formula of the compound	Calculated mass [M]	Experimental Mass [M-H] ⁻
16-Bzp-Biot	C ₃₃₇ H ₄₄₁ N ₁₀₃ O ₁₉₉ P ₃₀ S	100074.94	10078.19
16-Dzr-Biot	C ₃₃₂ H ₄₃₆ F ₃ N ₁₀₅ O ₁₉₈ P ₃₀ S	100078.90	10100.37
17-Iodo-Biot	C ₃₁₆ H ₄₁₇ IN ₁₀₂ O ₁₉₇ P ₃₀ S	9879.66	9895.42
H	C ₂₉₆ H ₃₆₇ N ₁₂₈ O ₁₇₂ P ₂₉ S	9366.62	9367.52
Ctrl-Biot	C ₃₁₇ H ₄₂₀ N ₁₀₂ O ₁₉₇ P ₃₀ S	9767.78	9774.56
16-Bzp	C ₃₁₀ H ₃₉₃ N ₉₈ O ₁₈₇ P ₂₉	9377.66	9379.97
17-Iodo	C ₂₈₉ H ₃₆₉ IN ₉₇ O ₁₈₅ P ₂₉	9182.38	9186.84

Table 9. Mass obtained by MALDI-TOF characterisation of oligodeoxynucleotides.

3) Radioactive labeling of DNA and duplex formation

[$\gamma^{32}\text{P}$]ATP (3000 Ci/mmol) was purchased from Perkin Elmer and Micro Bio-Spin® 6 columns from Biorad.

Labeling of ODNs

The indicated single-strand DNA was radiolabeled at its 5' end by T4 polynucleotide kinase and [$\gamma^{32}\text{P}$]ATP according to the manufacture instructions. Unincorporated [$\gamma^{32}\text{P}$]ATP was removed using a Biospin 6 column (Bio-Rad) equilibrated in TE buffer. Labeled oligonucleotides were checked for purity on a 15% Acrylamide 7M Urea denaturing gel and further purified if needed.

Preparation of duplex DNA

1.5 (15 nM) equivalents of the complementary non-labelled strand were added to the labeled oligodeoxynucleotide (10 nM) in buffer containing 20 mM Tris pH=7.5 and 10 mM LiCl. Samples were heated at 90°C for 5 minutes then allowed to cool back to room temperature, and stored at 4°C before use.

4) Duplex Stability

Electrophoretic Mobility Shift Assay (EMSA)

Increasing concentrations of the complementary non-labelled strand were added to the labeled oligodeoxynucleotide (10nM). Samples (10 µL) were heated at 90°C for 5 minutes then allowed to cool back to room temperature. Samples were then loaded on a 10% non-denaturing polyacrylamide gel (acrylamide:bisacrylamide 29:1) in 0.5X TBE buffer. Electrophoresis was performed for 45 min at room temperature. After electrophoresis, the gel was exposed to a phosphorimager screen. After being exposed for at least 10 h, the screen was scanned with the Phosphorimager TYPHOON instrument (Molecular Dynamics). The samples in the gel were quantified using ImageQuant version 5.1. Each experiment was reproduced at least three times. When shown, error bars correspond to the standard deviation calculated from at least three independent experiments.

Melting temperatures

Melting profiles were measured on a Perkin-Elmer λ40 spectrophotometer using quartz glass cuvettes with 1 cm path length. In these were in a total volume of 0.5 mL: 0.2 µM ODN, 0.2 µM counterstrand, 150 mM NaCl and 10 mM Tris buffer pH: 7.4. First, the oligonucleotides were hybridized by slowly cooling the samples down from 80°C to 20°C. The melting profiles started with a denaturing run (20 to 85°C) with a slope of 0.5 °C/min. One denaturing and one renaturing run were performed and averaged for evaluation of the melting point. For analysis of the data, the program Prism5 was used.

5) DNA/Protein interaction

Recombinant hRPA was expressed in Escherichia coli BL21 (DE3) strain transformed with the pET11ahRPA plasmid that permits the co-expression of RPA1, RPA2, and RPA3. hRPA

was purified over Affi-Gel Blue, Hydroxyapatite (Biorad), and Q-Sepharose chromatography columns as described previously²⁶³. The C-terminal catalytic domain of the murine Dnmt3a (623-908) was obtained as described in²⁶⁴.

Electrophoretic Mobility Shift Assay (EMSA)

For all EMSAs with hRPA, hRPA was diluted and preincubated (20 min at 4°C) in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA, and 0.1 mM EDTA. Radiolabeled oligonucleotide (10 nM) in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol) was incubated with various amounts of protein for 15 min at 20°C (10 µL). Samples were then loaded on a non-denaturing 5% acrylamide gel (acrylamide:bisacrylamide 29:1) in 0.5X TBE buffer. Electrophoresis was performed for 45 min at room temperature.

For EMSAs with Dnmt3a, duplexes (5 nM) were incubated in reaction buffer (20 mM HEPES pH=7.2, 1 mM EDTA, 50 mM KCl, 50 µg/mL BSA) with various amount of Dnmt3a and with or without 3 mM AdoMet. After 1h at 37°C, samples were loaded either on a non-denaturing 5% acrylamide gel (acrylamide:bisacrylamide 29:1) in 0.5X TBE buffer or on a non-denaturing 1% agarose gel. Electrophoresis was performed for 40 min at room temperature. Gel analysis was performed as previously described.

6) Expériences de photomarquage

Photocrosslinking kinetics with hRPA

hRPA was diluted and preincubated (20 min at 4°C) in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA, and 0.1 mM EDTA. RPA (7.5 equivalents, final concentration 75 nM) was added to the labeled oligodeoxynucleotide (final concentration 10 nM) in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol). Samples were incubated 30 min at 20°C and then irradiated for various amounts of time at 4°C with a 365 nm UV lamp (12W, Bioblock Scientific). Control samples were incubated at the same temperature but not irradiated. Samples were then loaded on a denaturing 10% polyacrylamide gel (375 mM Tris pH=8.8, 0.1% SDS). Electrophoresis was performed for 60 min at room temperature. Gel analysis was performed as previously described.

Photolabeling of DNMT3a

Duplexes (5 nM) were incubated in reaction buffer (20 mM HEPES pH=7.2, 1 mM EDTA, 50 mM KCl, 50 µg/mL BSA) with various amount of Dnmt3a and 3 mM AdoMet. After 1h at 37°C, samples were irradiated on ice with the UV lamp (12W, Bioblock Scientific) for 45 minutes. Control was incubated for 45 minutes on ice but not irradiated. Samples were then loaded on a denaturing 10% polyacrylamide gel (375 mM Tris pH=8.8, 0.1% SDS) and electrophoresis was performed for 45 minutes at room temperature. Gel analysis was performed as previously described.

Photolabeling with LEDs

hRPA was diluted and preincubated (20 min at 4°C) in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA, and 0.1 mM EDTA. RPA (7.5 equivalents, final concentration 75 nM) was added to the labeled oligodeoxynucleotide (final concentration 10 nM) in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol). Samples were incubated 30 min at 20°C and then irradiated for 45 minutes either with the UV lamp or the LEDs (300 or 365 nm). Samples were then loaded on a denaturing 10% polyacrylamide gel (375 mM Tris pH=8.8, 0.1% SDS) and electrophoresis was performed for 50 min at room temperature. Gel analysis was performed as previously described.

Photolabeling of bacterial lysates

To a solution containing 1 mg/mL of bacterial lysates (standard, overexpressing hRPA WT or overexpressing hRPA D228Y) in cell lysis buffer was added 1 nM of the labeled oligodeoxynucleotide. Control samples were prepared by adding purified RPA (8 nM final concentration), purified SSB (25 nM final concentration), or nothing to a solution of 1 nM labeled oligodeoxynucleotide in hRPA complexation buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol). Samples were incubated for 40 minutes at 4°C and irradiated for 30 minutes at 40°C with the UV lamp. Samples were then loaded on a denaturing 10% polyacrylamide gel (375 mM Tris pH=8.8, 0.1% SDS) and electrophoresis was performed for 70 min at room temperature. Gel analysis was performed as previously described.

Western Blot of protein lysates

Cell lysates from E.coli overexpressing hRPA or hRPA D228Y were loaded on a denaturing 10% acrylamide gel (acrylamide:bisacrylamide 19:1, 375 mM Tris-HCl pH=8.8, 0.1% SDS) in 1X TG-SDS Buffer. Total protein concentration was controlled using a Bradford assay. The protein species were transferred from the polyacrylamide gel to a PVDF membrane using the Trans-Blot® Turbo™ Transfer System. The membrane was washed in phosphate buffered saline, 0.1% Tween (PBST) buffer for 1 h at room temperature on an orbital shaker and then incubated in PBST buffer containing 5% non-fat milk at room temperature for 1 h on an orbital shaker. Finally, membranes were rinsed for 5 min with PBST. The standard ECL western blotting protocol from Amersham Biosciences was used to identify hRPA with antibodies specific for hRPA1 and hRPA2 subunits.

Magnetic beads purification and release

Labeled oligodeoxynucleotides in water were incubated with magnetic beads for 1 hour at room temperature on an orbital shaker. Using a magnet, the supernatant was then discarded and the beads were washed 2 times with buffer A (50 mM Tris pH=7.4; 0.1 mg/mL BSA), 2 times with buffer B (50 mM Tris pH=7.4 ; 0.1 mg/mL BSA ; 0.1% SDS) and 2 times with buffer C (50 mM Tris pH=7.4 ; 0.1 mg/mL BSA ; NaCl 1M). Oligodeoxynucleotides were then released from the beads with 15 minutes incubation at 95°C. Capture and release efficiency was assessed using a scintillation counter.

Photolabeling of RPA and purification with magnetic beads

hRPA was diluted and preincubated (20 min at 4°C) in buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol, 0.2 mg/mL BSA, and 0.1 mM EDTA. RPA (7.5 equivalents, final concentration 75 nM) was added to the labeled oligodeoxynucleotide (final concentration 10 nM) in reaction buffer (50 mM HEPES, pH 7.9, 0.1 mg/mL BSA, 100 mM KCl and 2% glycerol). Samples were incubated 20 min at 20°C and then irradiated for 60 minutes at 4°C with the UV lamp. Controls were not irradiated. Samples were then incubated with magnetic beads for 1 hour at room temperature on an orbital shaker. Using a magnet, the supernatant was then discarded and the beads were washed 2 times with buffer A (50 mM Tris pH=7.4; 0.1 mg/mL BSA), 2 times with buffer B (50 mM Tris pH=7.4 ; 0.1 mg/mL BSA ; 0.1% SDS) and 2 times with buffer C (50 mM Tris pH=7.4 ; 0.1 mg/mL BSA ; NaCl 1M). Oligodeoxynucleotides were then released from the beads with 5 minutes incubation at 95°C in NuPage 1X (gel loading buffer). Samples were then loaded on a denaturing 10%

polyacrylamide gel (375 mM Tris pH=8.8, 0.1% SDS) and electrophoresis was performed for 45 min at room temperature. Gel analysis was performed as previously described.

Photolabeling of nuclear extracts

Hela nuclear extracts were obtained from Santa Cruz Biotechnology. Rat brains were a gracious gift of Laurent Venance (Collège de France). Nuclear extracts from rat brain were prepared using the CellLytic NuCLEAR Extraction Kit (Sigma). Protein concentration was assessed using a Bradford assay. To a solution of the radiolabeled duplex (final concentration 5 nM) were added various amounts of nuclear extracts. Samples were incubated at 4°C for 45 minutes then irradiated at 4°C for 30 minutes with a UV lamp. Samples were then loaded on a denaturing 10% polyacrylamide gel (375 mM Tris pH=8.8, 0.1% SDS) and electrophoresis was performed for 45 minutes at room temperature. After electrophoresis, the gel was exposed to a phosphorimager screen. After being exposed for at least 10 h, the screen was scanned with the Phosphorimager TYPHOON instrument (Molecular Dynamics).

Bibliographie

1. Waddington, C. H. The Epigenotype. *Int. J. Epidemiol.* **41**, 10–13 (2012).
2. Jeltsch, A. Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem* **3**, 274–293 (2002).
3. Allis, C. D., Jenuwein, T. & Reinberg, D. *Epigenetics*. (Cold Spring Harbor Laboratory Press, 2007).
4. Goll, M. G. *et al.* Methylation of tRNA^{Asp} by the DNA Methyltransferase Homolog Dnmt2. *Science* **311**, 395–398 (2006).
5. Jurkowski, T. P. *et al.* Human DNMT2 methylates tRNA^{Asp} molecules using a DNA methyltransferase-like catalytic mechanism. *RNA* **14**, 1663–1670 (2008).
6. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
7. Illingworth, R. S. & Bird, A. P. CpG islands--'a rough guide'. *FEBS Lett.* **583**, 1713–1720 (2009).
8. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
9. Greger, V., Passarge, E., Hopping, W., Messmer, E. & Horsthemke, B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum. Genet.* **83**, 155–158 (1989).
10. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
11. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
12. Cohen, S. S. & Weed, L. L. Some precursors of the 5-hydroxymethylcytosine of T6r+ bacteriophage. *J. Biol. Chem.* **209**, 789–794 (1954).
13. Warren, R. A. J. Modified Bases in Bacteriophage DNAs. *Annu. Rev. Microbiol.* **34**, 137–158 (1980).

14. Revel, H. R. & Luria, S. E. DNA-Glucosylation in T-Even Phage: Genetic Determination and Role in Phage-Host Interaction. *Annu. Rev. Genet.* **4**, 177–192 (1970).
15. Penn, N. W., Suwalski, R., O’Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126**, 781–790 (1972).
16. Privat, E. & Sowers, L. C. Photochemical Deamination and Demethylation of 5-Methylcytosine. *Chem. Res. Toxicol.* **9**, 745–750 (1996).
17. Wagner, J. R. & Cadet, J. Oxidation Reactions of Cytosine DNA Components by Hydroxyl Radical and One-Electron Oxidants in Aerated Aqueous Solutions. *Acc. Chem. Res.* **43**, 564–571 (2010).
18. Madugundu, G. S., Cadet, J. & Wagner, J. R. Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA. *Nucleic Acids Res.* **42**, 7450–7460 (2014).
19. Cliffe, L. J. *et al.* JBP1 and JBP2 Proteins Are Fe²⁺/2-Oxoglutarate-dependent Dioxygenases Regulating Hydroxylation of Thymidine Residues in Trypanosome DNA. *J. Biol. Chem.* **287**, 19886–19895 (2012).
20. Bullard, W., Rosa-Spiegler, J. L. da, Liu, S., Wang, Y. & Sabatini, R. Identification of the Glucosyltransferase That Converts Hydroxymethyluracil to Base J in the Trypanosomatid Genome. *J. Biol. Chem.* **289**, 20273–20282 (2014).
21. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).
22. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
23. Mantri, M., Zhang, Z., McDonough, M. A. & Schofield, C. J. Autocatalysed oxidative modifications to 2-oxoglutarate dependent oxygenases. *FEBS J.* **279**, 1563–1575 (2012).

24. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
25. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
26. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
27. Fu, L. *et al.* Tet-Mediated Formation of 5-Hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136**, 11582–11585 (2014).
28. Dunwell, T. L., McGuffin, L. J., Dunwell, J. M. & Pfeifer, G. P. The mysterious presence of a 5-methylcytosine oxidase in the *Drosophila* genome. *Cell Cycle* **12**, 3357–3365 (2013).
29. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
30. Hu, L. *et al.* Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **155**, 1545–1555 (2013).
31. Hashimoto, H. *et al.* Structure of a *Naegleria* Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395 (2014).
32. Jin, C. *et al.* TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* **42**, 6956–6971 (2014).
33. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* **42**, 451–464 (2011).
34. Frauer, C. *et al.* Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1. *PLoS ONE* **6**, e16627 (2011).
35. Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122–126 (2013).

36. Xu, Y. *et al.* Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for *Xenopus* Eye and Neural Development. *Cell* **151**, 1200–1213 (2012).
37. Münzel, M. *et al.* Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry* **17**, 13782–13788 (2011).
38. Renciuik, D., Blacque, O., Vorlickova, M. & Spingler, B. Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.* **41**, 9891–9900 (2013).
39. Lercher, L. *et al.* Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **50**, 1794–1796 (2014).
40. Thalhammer, A., Hansen, A. S., El-Sagheer, A. H., Brown, T. & Schofield, C. J. Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun.* **47**, 5325–5327 (2011).
41. Maiti, A., Michelson, A. Z., Armwood, C. J., Lee, J. K. & Drohat, A. C. Divergent Mechanisms for Enzymatic Excision of 5-Formylcytosine and 5-Carboxylcytosine from DNA. *J. Am. Chem. Soc.* **135**, 15813–15822 (2013).
42. Shibutani, T. *et al.* Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci. Rep.* **4**, (2014).
43. Ji, D., You, C., Wang, P. & Wang, Y. Effects of Tet-Induced Oxidation Products of 5-Methylcytosine on DNA Replication in Mammalian Cells. *Chem. Res. Toxicol.* **27**, 1304–1309 (2014).
44. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).

45. Nestor, C. E. *et al.* Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res* **22**, 467–477 (2012).
46. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367–e15367 (2010).
47. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* **49**, 5375–5377 (2010).
48. Yang, H. *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**, 663–669 (2013).
49. Haffner, M. C. *et al.* Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* **2**, 627–637 (2011).
50. Kudo, Y. *et al.* Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci* **103**, 670–676 (2012).
51. Liu, S. *et al.* Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res.* **41**, 6421–6429 (2013).
52. Jin, S.-G. *et al.* 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res.* **71**, 7360–7365 (2011).
53. Pfaffeneder, T. *et al.* The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem. Int. Ed.* **50**, 7008–7012 (2011).
54. Zhang, L. *et al.* A TET Homologue Protein from *Coprinopsis cinerea* (CcTET) That Biochemically Converts 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine. *J. Am. Chem. Soc.* **136**, 4801–4804 (2014).
55. Szulwach, K. E. *et al.* 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci* **14**, 1607–1616 (2011).

56. Ficiz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
57. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* **25**, 2436–2452 (2011).
58. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* **151**, 1417–1430 (2012).
59. Jin, S.-G., Wu, X., Li, A. X. & Pfeifer, G. P. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res* **39**, 5015–5024 (2011).
60. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25**, 679–684 (2011).
61. Booth, M. J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).
62. Yu, M. *et al.* Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* **149**, 1368–1380 (2012).
63. Booth, M. J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **6**, 435–440 (2014).
64. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **advance online publication**, (2014).
65. Almeida, R. D. *et al.* Semi-quantitative immunohistochemical detection of 5-hydroxymethyl-cytosine reveals conservation of its tissue distribution between amphibians and mammals. *Epigenetics* **7**, 137–140 (2012).

66. Almeida, R. D. *et al.* 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development. *Epigenetics* **7**, 383–389 (2012).
67. Bracht, J. R., Perlman, D. H. & Landweber, L. F. Cytosine methylation and hydroxymethylation mark DNA for elimination in *Oxytricha trifallax*. *Genome Biol.* **13**, R99 (2012).
68. Wojciechowski, M. *et al.* Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biol.* **4**, 140110 (2014).
69. Liu, S. *et al.* Detection of Oxidation Products of 5-Methyl-2'-Deoxycytidine in Arabidopsis DNA. *PLoS ONE* **8**, e84620 (2013).
70. Valinluck, V. & Sowers, L. C. Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).
71. Otani, J. *et al.* Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells. *PLoS ONE* **8**, e82961 (2013).
72. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* (2012). doi:10.1093/nar/gks155
73. Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* **6**, e21306–e21306 (2011).
74. Otani, J. *et al.* Structural Basis of the Versatile DNA Recognition Ability of the Methyl-CpG Binding Domain of Methyl-CpG Binding Domain Protein 4. *J. Biol. Chem.* **288**, 6351–6362 (2013).
75. Ji, D., Lin, K., Song, J. & Wang, Y. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. Biosyst.* **10**, 1749–1752 (2014).

76. Kriukienė, E., Liutkevičiūtė, Z. & Klimašauskas, S. 5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA. *Chem. Soc. Rev.* **41**, 6916–6930 (2012).
77. Wu, H. & Zhang, Y. Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell* **156**, 45–68 (2014).
78. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *J. Biol. Chem.* **286**, 35334–35338 (2011).
79. Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X. & Cheng, X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **40**, 10203–10214 (2012).
80. Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
81. Boorstein, R. J. *et al.* Definitive Identification of Mammalian 5-Hydroxymethyluracil DNAN-Glycosylase Activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).
82. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
83. Kumar, R. *et al.* AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* **500**, 89–92 (2013).
84. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).
85. Jang, H., Shin, H., Eichman, B. F. & Huh, J. H. Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. *Biochem. Biophys. Res. Commun.* **446**, 1067–1072 (2014).

86. Schiesser, S. *et al.* Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing. *Angew. Chem. Int. Ed.* **51**, 6516–6520 (2012).
87. Schiesser, S. *et al.* Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
88. Simmons, J. M., Müller, T. A. & Hausinger, R. P. FeII/ α -ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism. *Dalton Trans.* 5132–5142 (2008). doi:10.1039/B803512A
89. Xu, S. *et al.* Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res.* **23**, 1296–1309 (2013).
90. Liutkevičiūtė, Z., Lukinavičius, G., Masevičius, V., Daujotytė, D. & Klimašauskas, S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nat. Chem. Biol.* **5**, 400–402 (2009).
91. Chen, C.-C., Wang, K.-Y. & Shen, C.-K. J. The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B Are Also DNA 5-Hydroxymethylcytosine Dehydroxymethylases. *J. Biol. Chem.* **287**, 33116–33121 (2012).
92. Chen, C.-C., Wang, K.-Y. & Shen, C.-K. J. DNA 5-Methylcytosine Demethylation Activities of the Mammalian DNA Methyltransferases. *J. Biol. Chem.* **288**, 9084–9091 (2013).
93. Liutkevičiūtė, Z. *et al.* Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
94. Hashimoto, H. *et al.* Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev.* **28**, 2304–2313 (2014).

95. Liu, Y., Toh, H., Sasaki, H., Zhang, X. & Cheng, X. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev.* **26**, 2374–2379 (2012).
96. Liu, Y., Olanrewaju, Y. O., Zhang, X. & Cheng, X. DNA Recognition of 5-Carboxylecytosine by a Zfp57 Mutant at an Atomic Resolution of 0.97 Å. *Biochemistry (Mosc.)* **52**, 9310–9317 (2013).
97. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell* **153**, 480–492 (2013).
98. Yildirim, O. *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* **147**, 1498–1510 (2011).
99. Lafaye, C. *et al.* DNA binding of the p21 repressor ZBTB2 is inhibited by cytosine hydroxymethylation. *Biochem. Biophys. Res. Commun.* **446**, 341–346 (2014).
100. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152**, 1146–1159 (2013).
101. Dejosez, M. *et al.* Ronin Is Essential for Embryogenesis and the Pluripotency of Mouse Embryonic Stem Cells. *Cell* **133**, 1162–1174 (2008).
102. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Embryogenesis: Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
103. Gu, T.-P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
104. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res* **21**, 1670–1676 (2011).

105. Koh, K. P. *et al.* Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. *Cell Stem Cell* **8**, 200–213 (2011).
106. Hackett, J. A. *et al.* Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science* **339**, 448–452 (2013).
107. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
108. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
109. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–374 (2013).
110. Zhang, P., Huang, B., Xu, X. & Sessa, W. C. Ten-eleven translocation (Tet) and thymine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a. *Biochem. Biophys. Res. Commun.* **437**, 368–373 (2013).
111. Morita, S. *et al.* miR-29 Represses the Activities of DNA Methyltransferases and DNA Demethylases. *Int. J. Mol. Sci.* **14**, 14647–14658 (2013).
112. Cheng, J. *et al.* An Extensive Network of TET2-Targeting MicroRNAs Regulates Malignant Hematopoiesis. *Cell Rep.* **5**, 471–481 (2013).
113. Fu, X. *et al.* MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation. *Proc. Natl. Acad. Sci.* **110**, 17892–17897 (2013).
114. Vella, P. *et al.* Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Mol. Cell* **49**, 645–656 (2013).
115. Ito, R. *et al.* TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* **19**, 52–65 (2014).

116. Wang, Y. & Zhang, Y. Regulation of TET Protein Stability by Calpains. *Cell Rep.* **6**, 278–284 (2014).
117. Lorschach, R. B. *et al.* TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **17**, 637–641 (2003).
118. Ono, R. *et al.* LCX, Leukemia-associated Protein with a CXXC Domain, Is Fused to MLL in Acute Myeloid Leukemia with Trilineage Dysplasia Having t(10;11)(q22;q23). *Cancer Res.* **62**, 4075–4080 (2002).
119. Huang, H. *et al.* TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci.* **110**, 11994–11999 (2013).
120. Hsu, C.-H. *et al.* TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. *Cell Rep.* **2**, 568–579 (2012).
121. Delhommeau, F. *et al.* Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
122. Kosmider, O. *et al.* TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs). *Blood* **114**, 3285–3291 (2009).
123. Abdel-Wahab, O. *et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144–147 (2009).
124. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
125. Pérez, C. *et al.* TET2 Mutations Are Associated with Specific 5-Methylcytosine and 5-Hydroxymethylcytosine Profiles in Patients with Chronic Myelomonocytic Leukemia. *PLoS ONE* **7**, e31605 (2012).
126. Ko, M. *et al.* Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U A* **108**, 14566–14571 (2011).

127. Abdel-Wahab, O. & Levine, R. L. Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood* **121**, 3563–3572 (2013).
128. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
129. Xu, W. *et al.* Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α -Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
130. Kats, L. M. *et al.* Proto-Oncogenic Role of Mutant IDH2 in Leukemia Initiation and Maintenance. *Cell Stem Cell* **14**, 329–341 (2014).
131. Sasaki, M. *et al.* IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488**, 656–659 (2012).
132. Xiao, M. *et al.* Inhibition of -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).
133. Dong, E., Gavin, D. P., Chen, Y. & Davis, J. Upregulation of TET1 and downregulation of APOBEC3A and APOBEC3C in the parietal cortex of psychotic patients. *Transl. Psychiatry* **2**, e159 (2012).
134. International Statistical Classification of Diseases and Related Health Problems 10th Revision (F84). at <<http://apps.who.int/classifications/icd10/browse/2010/en>>
135. Zhubi, A. *et al.* Increased binding of MeCP2 to the GAD1 and RELN promoters may be mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Transl. Psychiatry* **4**, e349 (2014).
136. Coppieters, N. *et al.* Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* **35**, 1334–1344 (2014).
137. Bradley-Whitman, M. A. & Lovell, M. A. Epigenetic changes in the progression of Alzheimer's disease. *Mech. Ageing Dev.* **134**, 486–495 (2013).

138. Gowher, H., Ehrlich, K. C. & Jeltsch, A. DNA from *Aspergillus flavus* contains 5-methylcytosine. *FEMS Microbiol. Lett.* **205**, 151–155 (2001).
139. Fisher, D. H. & Giese, R. W. Determination of 5-methylcytosine in DNA by gas chromatography-electron-capture detection. *J. Chromatogr.* **452**, 51–60 (1988).
140. Ramsahoye, B. H. Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. *Methods San Diego Calif* **27**, 156–161 (2002).
141. Fraga, M. F. *et al.* High-performance capillary electrophoretic method for the quantification of 5-methyl 2'-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. *Electrophoresis* **23**, 1677–1681 (2002).
142. Song, L., James, S. R., Kazim, L. & Karpf, A. R. Specific Method for the Determination of Genomic DNA Methylation by Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry. *Anal. Chem.* **77**, 504–510 (2005).
143. Guianvarc'h, D., Arimondo, P. B., Blasco, T. & Bolbach, G. Analyse par LC-ESI-MS/MS de la quantité de 5-méthyl-2'-désoxycytidine dans l'ADN génomique: limite de détection et domaine de linéarité. in (2008).
144. Champion, C. *et al.* Mechanistic Insights on the Inhibition of C5 DNA Methyltransferases by Zebularine. *PLoS ONE* **5**, e12388 (2010).
145. Desjobert, C. *et al.* Combined analysis of DNA methylation and cell cycle in cancer cells. *Epigenetics Off. J. DNA Methylation Soc.* **0** (2014).
doi:10.1080/15592294.2014.995542
146. Thomson, J. P. *et al.* Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Res.* **41**, e206–e206 (2013).
147. Matarese, F., Pau, E. C. S. & Stunnenberg, H. G. 5-Hydroxymethylcytosine: a new kid on the epigenetic block? *Mol. Syst. Biol.* **7**, 562 (2011).

148. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
149. Song, C.-X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* **29**, 68–72 (2011).
150. Song, C.-X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
151. Robertson, A. B., Dahl, J. A., Ougland, R. & Klungland, A. Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nat. Protoc.* **7**, 340–350 (2012).
152. Robertson, A. B. *et al.* A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **39**, e55–e55 (2011).
153. Li, W.-W., Gong, L. & Bayley, H. Single-Molecule Detection of 5-Hydroxymethylcytosine in DNA through Chemical Modification and Nanopore Analysis. *Angew. Chem. Int. Ed.* **52**, 4350–4355 (2013).
154. Raiber, E.-A. *et al.* Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* **13**, R69 (2012).
155. Szwagierczak, A. *et al.* Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.* **39**, 5149–5156 (2011).
156. Cohen-Karni, D. *et al.* The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci.* **108**, 11040–11045 (2011).
157. Bhattacharyya, S. *et al.* Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. *Nucleic Acids Res.* **41**, e157–e157 (2013).

158. Wang, H. *et al.* Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.* **39**, 9294–9305 (2011).
159. Sun, Z. *et al.* High-Resolution Enzymatic Mapping of Genomic 5-Hydroxymethylcytosine in Mouse Embryonic Stem Cells. *Cell Rep.* **3**, 567–576 (2013).
160. Kriukienė, E., Liutkevičiūtė, Z. & Klimašauskas, S. 5-Hydroxymethylcytosine - the elusive epigenetic mark in mammalian DNA. *Chem. Soc. Rev.* **41**, 6916–6930 (2012).
161. Huang, Y. *et al.* The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS ONE* **5**, e8888 (2010).
162. Nestor, C., Ruzov, A., Meehan, R. & Dunican, D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques* **48**, 317–319 (2010).
163. Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* **7**, 2159–2170 (2012).
164. Lu, X. *et al.* Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA. *J. Am. Chem. Soc.* **135**, 9315–9317 (2013).
165. Ning, X., Guo, J., Wolfert, M. A. & Boons, G.-J. Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angew. Chem. Int. Ed.* **47**, 2253–2255 (2008).
166. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
167. Wallace, E. V. B. *et al.* Identification of epigenetic DNA modifications with a protein nanopore. *Chem. Commun.* **46**, 8195–8197 (2009).
168. Wanunu, M. *et al.* Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules. *J. Am. Chem. Soc.* **133**, 486–492 (2011).

169. Manrao, E. A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
170. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc. Natl. Acad. Sci.* **110**, 18904–18909 (2013).
171. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**, 133–138 (2009).
172. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
173. Song, C.-X. *et al.* Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods* **9**, 75–77 (2012).
174. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
175. Larivière, L. & Moréra, S. A base-flipping mechanism for the T4 phage beta-glucosyltransferase and identification of a transition-state analog. *J Mol Biol* **324**, 483–490 (2002).
176. Larivière, L., Gueguen-Chaignon, V. & Moréra, S. Crystal structures of the T4 phage beta-glucosyltransferase and the D100A mutant in complex with UDP-glucose: glucose binding and identification of the catalytic base for a direct displacement mechanism. *J. Mol. Biol.* **330**, 1077–1086 (2003).
177. Larivière, L., Kurzeck, J., Aschke-Sonnenborn, U., Rüger, W. & Moréra, S. Crystallization and preliminary crystallographic study of a ternary complex between the T4 phage beta-glucosyltransferase, uridine diphosphoglucose and a DNA fragment containing an abasic site. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1484–1486 (2002).

178. Moréra, S. *et al.* High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding. *J. Mol. Biol.* **311**, 569–577 (2001).
179. Münzel, M., Globisch, D., Trindler, C. & Carell, T. Efficient synthesis of 5-hydroxymethylcytosine containing DNA. *Org. Lett.* **12**, 5671–5673 (2010).
180. Wagner, G. K., Pesnot, T. & Field, R. A. A survey of chemical methods for sugar-nucleotide synthesis. *Nat. Prod. Rep.* **26**, 1172–1194 (2009).
181. Bosco, M. *et al.* 6-Azido d-galactose transfer to N-acetyl-d-glucosamine derivative using commercially available β -1,4-galactosyltransferase. *Tetrahedron Lett.* **49**, 2294–2297 (2008).
182. Shahal, T. *et al.* Spectroscopic Quantification of 5-Hydroxymethylcytosine in Genomic DNA. *Anal. Chem.* **86**, 8231–8237 (2014).
183. Hong, H. & Wang, Y. Derivatization with Girard reagent T combined with LC-MS/MS for the sensitive detection of 5-formyl-2'-deoxyuridine in cellular DNA. *Anal. Chem.* **79**, 322–326 (2007).
184. Sugizaki, K., Ikeda, S., Yanagisawa, H. & Okamoto, A. Facile synthesis of hydroxymethylcytosine-containing oligonucleotides and their reactivity upon osmium oxidation. *Org. Biomol. Chem.* **9**, 4176–4181 (2011).
185. El Safadi, Y. *et al.* 5-Modified-2'-dU and 2'-dC as Mutagenic Anti HIV-1 Proliferation Agents: Synthesis and Activity. *J. Med. Chem.* **53**, 1534–1545 (2010).
186. Lusic, H. *et al.* Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNAMet. *Nucleic Acids Res.* **36**, 6548–6557 (2008).

187. Cosner, C. C., Cabrera, P. J., Byrd, K. M., Thomas, A. M. A. & Helquist, P. Selective Oxidation of Benzylic and Allylic Alcohols Using Mn(OAc)₃/Catalytic 2,3-Dichloro-5,6-dicyano-1,4-benzoquinone. *Org. Lett.* **13**, 2071–2073 (2011).
188. Tang, Y. *et al.* Hydrophilic Material for the Selective Enrichment of 5-Hydroxymethylcytosine and Its Liquid Chromatography–Tandem Mass Spectrometry Detection. *Anal. Chem.* **85**, 6129–6135 (2013).
189. Guo, P. *et al.* Selective Detection of 5-Formyl-2'-deoxycytidine in DNA Using a Fluorogenic Hydroxylamine Reagent. *Org. Lett.* **15**, 3266–3269 (2013).
190. Hu, J. *et al.* Discrimination between 5-hydroxymethylcytosine and 5-methylcytosine in DNA by selective chemical labeling. *Bioorg. Med. Chem. Lett.* **24**, 294–297 (2014).
191. Hong, T. *et al.* Fluorescent Strategy Based on Cationic Conjugated Polymer Fluorescence Resonance Energy Transfer for the Quantification of 5-(Hydroxymethyl)cytosine in Genomic DNA. *Anal. Chem.* **85**, 10797–10802 (2013).
192. Singh, A., Thornton, E. R. & Westheimer, F. H. The photolysis of diazoacetylchymotrypsin. *J. Biol. Chem.* **237**, 3006–3008 (1962).
193. Dormán, G. & Prestwich, G. D. Using photolabile ligands in drug discovery and development. *Trends Biotechnol.* **18**, 64–77 (2000).
194. Vodovozova, E. L. Photoaffinity labeling and its application in structural biology. **72**, 1–20 (2007).
195. Yip, C. C., Yeung, C. W. & Moule, M. L. Photoaffinity labeling of insulin receptor of rat adipocyte plasma membrane. *J. Biol. Chem.* **253**, 1743–1745 (1978).
196. Eirich, J., Orth, R. & Sieber, S. A. Unraveling the Protein Targets of Vancomycin in Living *S. aureus* and *E. faecalis* Cells. *J. Am. Chem. Soc.* **133**, 12144–12153 (2011).

197. Chakrabarti, P. & Khorana, H. G. New approach to the study of phospholipid-protein interactions in biological membranes. Synthesis of fatty acids and phospholipids containing photosensitive groups. *Biochemistry (Mosc.)* **14**, 5021–5033 (1975).
198. Palestini, P. *et al.* Tubulin Anchoring to Glycolipid-enriched, Detergent-resistant Domains of the Neuronal Plasma Membrane. *J. Biol. Chem.* **275**, 9978–9985 (2000).
199. Fleet, G. W. J., Porter, R. R. & Knowles, J. R. Affinity Labelling of Antibodies with Aryl Nitrene as Reactive Group. *Nature* **224**, 511–512 (1969).
200. Kotzyba-Hibert, F., Kapfer, I. & Goeldner, M. Recent Trends in Photoaffinity Labeling. *Angew. Chem. Int. Ed. Engl.* **34**, 1296–1312 (1995).
201. Leyva, E., Platz, M. S., Persy, G. & Wirz, J. Photochemistry of phenyl azide: the role of singlet and triplet phenylnitrene as transient intermediates. *J. Am. Chem. Soc.* **108**, 3783–3790 (1986).
202. Platz, M. S. Comparison of Phenylcarbene and Phenylnitrene. *Acc. Chem. Res.* **28**, 487–492 (1995).
203. Poe, R., Schnapp, K., Young, M. J. T., Grayzar, J. & Platz, M. S. Chemistry and kinetics of singlet pentafluorophenylnitrene. *J. Am. Chem. Soc.* **114**, 5054–5067 (1992).
204. A. Fleming, S. Chemical reagents in photoaffinity labeling. *Tetrahedron* **51**, 12479–12520 (1995).
205. Galardy, R. E., Craig, L. C., Jamieson, J. D. & Printz, M. P. Photoaffinity Labeling of Peptide Hormone Binding Sites. *J. Biol. Chem.* **249**, 3510–3518 (1974).
206. Boyd, N. D., Kage, R., Dumas, J. J., Krause, J. E. & Leeman, S. E. The peptide binding site of the substance P (NK-1) receptor localized by a photoreactive analogue of substance P: presence of a disulfide bond. *Proc. Natl. Acad. Sci.* **93**, 433–437 (1996).
207. Kage, R., Leeman, S. E., Krause, J. E., Costello, C. E. & Boyd, N. D. Identification of Methionine as the Site of Covalent Attachment of a p-Benzoyl-Phenylalanine-containing

- Analogue of Substance P on the Substance P (NK-1) Receptor. *J. Biol. Chem.* **271**, 25797–25800 (1996).
208. Dorman, G. & Prestwich, G. D. Benzophenone Photophores in Biochemistry. *Biochemistry (Mosc.)* **33**, 5661–5673 (1994).
 209. Wittelsberger, A., Thomas, B. E., Mierke, D. F. & Rosenblatt, M. Methionine acts as a ‘magnet’ in photoaffinity crosslinking experiments. *FEBS Lett.* **580**, 1872–1876 (2006).
 210. Smith, R. A. G. & Knowles, J. R. Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *J. Am. Chem. Soc.* **95**, 5072–5073 (1973).
 211. Smith, R. A. G. & Knowles, J. R. The preparation and photolysis of 3-aryl-3H-diazirines. *J. Chem. Soc. Perkin Trans. 2* 686–694 (1975). doi:10.1039/P29750000686
 212. Brunner, J., Senn, H. & Richards, F. M. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J. Biol. Chem.* **255**, 3313–3318 (1980).
 213. Platz, M. *et al.* Photolysis of 3-aryl-3-(trifluoromethyl)diazirines: a caveat regarding their use in photoaffinity probes. *Bioconjug. Chem.* **2**, 337–341 (1991).
 214. Dubinsky, L., Krom, B. P. & Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg. Med. Chem.* **20**, 554–570 (2012).
 215. Shugar, D. & Fox, J. J. Absorption spectra and structure of 2-thiouracil derivatives as a function of pH. *Bull. Sociétés Chim. Belg.* **61**, 293–309 (2010).
 216. Gauss, D. H. & Sprinzl, M. Compilation of tRNA sequences. *Nucleic Acids Res.* **9**, r1–r23 (1981).
 217. Yaniv, M., Favre, A. & Barrell, B. G. Structure of Transfer RNA: Evidence for Interaction between Two Non-adjacent Nucleotide Residues in tRNA^{Val} from *Escherichia coli*. *Nature* **223**, 1331–1333 (1969).

218. Zhang, X. & Xu, Y.-Z. NMR and UV Studies of 4-Thio-2'-deoxyuridine and Its Derivatives. *Molecules* **16**, 5655–5664 (2011).
219. Reichardt, C. & Crespo-Hernández, C. E. Room-Temperature Phosphorescence of the DNA Monomer Analogue 4-Thiothymidine in Aqueous Solutions after UVA Excitation. *J. Phys. Chem. Lett.* **1**, 2239–2243 (2010).
220. Favre, A., Saintomé, C., Fourrey, J. L., Clivio, P. & Laugâa, P. Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid-protein interactions. *J. Photochem. Photobiol. B* **42**, 109–124 (1998).
221. Ito, S., Saito, I., Nakata, A. & Matsuura, T. Photochemical synthesis of model compounds for nucleic acid-protein adducts. *Nucleic Acids Res.* **1**, s321–s326 (1978).
222. Saenger, W. *Principles of nucleic acid structure*. (Springer-Verlag, 1984). at <http://catalog.hathitrust.org/api/volumes/oclc/9016993.html>
223. Sintim, H. O. & Kool, E. T. Enhanced Base Pairing and Replication Efficiency of Thiothymidines, Expanded-size Variants of Thymidine. *J. Am. Chem. Soc.* **128**, 396–397 (2006).
224. Somerville, L. *et al.* Structure and Dynamics of Thioguanine-modified Duplex DNA. *J. Biol. Chem.* **278**, 1005–1011 (2003).
225. Bokemeier, W. & Lezius, A. G. On the structure of poly d(A-S4T). *Nucleic Acids Res.* **14**, 2241–2249 (1986).
226. Ogata, R. & Gilbert, W. Contacts between the lac repressor and the thymines in the lac operator. *Proc. Natl. Acad. Sci.* **74**, 4973–4976 (1977).
227. Stump, W. T. & Hall, K. B. Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA* **1**, 55–63 (1995).

228. Norris, C. L., Meisenheimer, K. M. & Koch, T. H. Mechanistic Studies Relevant to Bromouridine-Enhanced Nucleoprotein Photocrosslinking: Possible Involvement of an Excited Tyrosine Residue of the Protein. *Photochem. Photobiol.* **65**, 201–207 (1997).
229. Norris, C. L., Meisenheimer, P. L. & Koch, T. H. Mechanistic Studies of the 5-Iodouracil Chromophore Relevant to Its Use in Nucleoprotein Photo-Cross-Linking. *J. Am. Chem. Soc.* **118**, 5796–5803 (1996).
230. Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R. & Koch, T. H. Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science* **262**, 1255–1257 (1993).
231. Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*. (Cornell University Press, 1960).
232. Lavrik, O. I. Photoaffinity Labeling of Mouse Fibroblast Enzymes by a Base Excision Repair Intermediate. EVIDENCE FOR THE ROLE OF POLY(ADP-RIBOSE) POLYMERASE-1 IN DNA REPAIR. *J. Biol. Chem.* **276**, 25541–25548 (2001).
233. Wlassoff, W. A. *et al.* Synthesis and Characterization of (d)NTP Derivatives Substituted with Residues of Different Photoreagents. *Bioconjug. Chem.* **6**, 352–360 (1995).
234. Belousova, E. A. *et al.* Photoreactive DNA probes as a tool for studying the translesion synthesis system in Mammalian cell extracts. *Med Chem* **4**, 155–162 (2008).
235. Gramlich, P. M. E., Wirges, C. T., Manetto, A. & Carell, T. Postsynthetic DNA Modification through the Copper-Catalyzed Azide-Alkyne Cycloaddition Reaction. *Angew. Chem. Int. Ed.* **47**, 8350–8358 (2008).

236. Connolly, B. A. & Newman, P. C. Synthesis and properties of oligonucleotides containing 4-thiothymidine, 5-methyl-2-pyrimidinone-1- β -D(2'-dexoyriboside) and 2-thiothymidine. *Nucleic Acids Res.* **17**, 4957–4974 (1989).
237. Shah, K., Wu, H. & Rana, T. M. Synthesis of Uridine Phosphoramidite Analogs: Reagents for Site-Specific Incorporation of Photoreactive Sites into RNA Sequences. *Bioconjug. Chem.* **5**, 508–512 (1994).
238. Winnacker, M., Welzmler, V., Strasser, R. & Carell, T. Development of a DNA photoaffinity probe for the analysis of 8-OxodG-binding proteins in a human proteome. *Chembiochem* **11**, 1345–1349 (2010).
239. Liebmann, M., Di Pasquale, F. & Marx, A. A new photoactive building block for investigation of DNA backbone interactions: photoaffinity labeling of human DNA polymerase beta. *Chembiochem* **7**, 1965–1969 (2006).
240. Marx, A. *et al.* Synthesis of 4'-C-Acylated Thymidines. *Helv. Chim. Acta* **79**, 1980–1994 (1996).
241. Weisbrod, S. H. & Marx, A. Novel strategies for the site-specific covalent labelling of nucleic acids. *Chem. Commun. Camb. Engl.* 5675–5685 (2008). doi:10.1039/b809528k
242. MacMillan, A. M. & Verdine, G. L. Synthesis of functionally tethered oligodeoxynucleotides by the convertible nucleoside approach. *J. Org. Chem.* **55**, 5931–5933 (1990).
243. Macmillan, A. M. & Verdine, G. L. Engineering tethered DNA molecules by the convertible nucleoside approach. *Tetrahedron* **47**, 2603–2616 (1991).
244. Erlanson, D. A., Chen, L. & Verdine, G. L. DNA methylation through a locally unpaired intermediate. *J. Am. Chem. Soc.* **115**, 12583–12584 (1993).
245. Caprioara, M., Fiammengio, R., Engeser, M. & Jäschke, A. DNA-based phosphane ligands. *Chemistry* **13**, 2089–2095 (2007).

246. Harris, C. M., Zhou, L., Strand, E. A. & Harris, T. M. New strategy for the synthesis of oligodeoxynucleotides bearing adducts at exocyclic amino sites of purine nucleosides. *J. Am. Chem. Soc.* **113**, 4328–4329 (1991).
247. Wang, H. *et al.* Site-specific synthesis of oligonucleotides containing malondialdehyde adducts of deoxyguanosine and deoxyadenosine via a postsynthetic modification strategy. *Chem Res Toxicol* **19**, 1467–1474 (2006).
248. Allerson, C. R., Chen, S. L. & Verdine, G. L. A Chemical Method for Site-Specific Modification of RNA: The Convertible Nucleoside Approach. *J. Am. Chem. Soc.* **119**, 7423–7433 (1997).
249. De la Torre, B. G. *et al.* Synthesis of Oligonucleotides Carrying Anchoring Groups and Their Use in the Preparation of Oligonucleotide–Gold Conjugates. *Helv. Chim. Acta* **85**, 2594–2607 (2002).
250. Xu, Y. Z., Zheng, Q. & Swann, P. F. Synthesis of DNA containing modified bases by post-synthetic substitution. Synthesis of oligomers containing 4-substituted thymine: O4-alkylthymine, 5-methylcytosine, N4-dimethylamino-5-methylcytosine, and 4-thiothymine. *J. Org. Chem.* **57**, 3839–3845 (1992).
251. Haginoya, N., Ono, A., Nomura, Y., Ueno, Y. & Matsuda, A. Nucleosides and Nucleotides. 160. Synthesis of Oligodeoxyribonucleotides Containing 5-(N-Aminoalkyl)carbamoyl-2'-deoxyuridines by a New Postsynthetic Modification Method and Their Thermal Stability and Nuclease-Resistance Properties†. *Bioconjug. Chem.* **8**, 271–280 (1997).
252. Shigdel, U. K., Zhang, J. & He, C. Diazirine-based DNA photo-cross-linking probes for the study of protein-DNA interactions. *Angew Chem Int Ed Engl* **47**, 90–93 (2008).

253. Coleman, R. S. & Kesicki, E. A. Synthesis and Postsynthetic Modification of Oligodeoxynucleotides Containing 4-Thio-2'-deoxyuridine (ds4U). *J. Am. Chem. Soc.* **116**, 11636–11642 (1994).
254. Coleman, R. S. & Siedlecki, J. M. Synthesis of a 4-thio-2'-deoxyuridine containing oligonucleotide. Development of the thiocarbonyl group as a linker element. *J. Am. Chem. Soc.* **114**, 9229–9230 (1992).
255. Coleman, R. S. & Pires, R. M. Covalent cross-linking of duplex DNA using 4-thio-2'-deoxyuridine as a readily modifiable platform for introduction of reactive functionality into oligonucleotides. *Nucleic Acids Res.* **25**, 4771–4777 (1997).
256. Lercher, L., McGouran, J. F., Kessler, B. M., Schofield, C. J. & Davis, B. G. DNA Modification under Mild Conditions by Suzuki–Miyaura Cross-Coupling for the Generation of Functional Probes. *Angew. Chem. Int. Ed.* **52**, 10553–10558 (2013).
257. Okamoto, A., Taiji, T., Tainaka, K. & Saito, I. Oligonucleotides containing 7-vinyl-7-deazaguanine as a facile strategy for expanding the functional diversity of DNA. *Bioorg. Med. Chem. Lett.* **12**, 1895–1896 (2002).
258. Fairman, M. P. & Stillman, B. Cellular factors required for multiple stages of SV40 DNA replication in vitro. *EMBO J.* **7**, 1211 (1988).
259. Kasama, T., Ono, A. & Izuta, S. Binding activity of replication protein A to single-stranded DNA containing oxidized pyrimidine base. *Nucleic Acids Symp. Ser.* **1**, 167–168 (2001).
260. Irie, D., Ono, A. & Izuta, S. Recognition of Oxidized Thymine Base on the Single-Stranded DNA by Replication Protein A. *Nucleosides Nucleotides Nucleic Acids* **25**, 439–451 (2006).

261. Rosenfeld, K. K., Ziv, T., Goldin, S., Glaser, F. & Manor, H. Mapping of DNA Binding Sites in the Tetrahymena Telomerase Holoenzyme Proteins by UV Cross-Linking and Mass Spectrometry. *J. Mol. Biol.* **410**, 77–92 (2011).
262. Burlina, F., Sagan, S., Bolbach, G. & Chassaing, G. A direct approach to quantification of the cellular uptake of cell-penetrating peptides using MALDI-TOF mass spectrometry. *Nat. Protoc.* **1**, 200–205 (2006).
263. Gomes, X. V., Henricksen, L. A. & Wold, M. S. Proteolytic Mapping of Human Replication Protein A: Evidence for Multiple Structural Domains and a Conformational Change upon Interaction with Single-Stranded DNA†. *Biochemistry (Mosc.)* **35**, 5586–5595 (1996).
264. Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* **277**, 20409–20414 (2002).

Table des illustrations

Figure 1. La réaction de méthylation de l'ADN par les DNMT.	14
Figure 2. Mécanisme de la méthylation des cytosines catalysé par les DNMTs.	14
Figure 3. La méthylation de novo par les DNMT3 et la méthylation de maintenance par DNMT1 dans l'ADN ³ .	15
Figure 4. La régulation de l'expression des gènes est étroitement liée à la méthylation de l'ADN.	16
Figure 5. Biosynthèse de la 5-hydroxyméthylcytosine, la 6 ^{ème} base de l'ADN.	17
Figure 6. La réaction catalysée par la dCMP hydroxyméthylase permettant la conversion de dCMP en d5hmCMP chez le phage T4.	18
Figure 7. Glucosylation des 5-hydroxyméthylcytosines par les enzymes AGT ou BGT chez le phage T4.	19
Figure 8. Formation de dérivés modifiés de la 5mC consécutivement à des dégâts oxydatifs.	20
Figure 9. Structure de la base J retrouvée exclusivement dans l'ADN des Trypanosomatidae.	21
Figure 10. Formation de la base J dans l'ADN des Trypanosomatidae.	21
Figure 11. La famille des protéines TET chez les mammifères. A) Séquence schématique des enzymes TETs humaines. B) Evolution des enzymes TET chez les mammifères ²² .	23
Figure 12. Cycle catalytique caractéristique des dioxygénases à Fe(II) dépendantes du 2-oxoglutarate ²³ .	24
Figure 13. Oxydations successives de la 5mC par les enzymes TET.	25
Figure 14. Poche catalytique de l'enzyme TET2, en complexe avec un duplex de l'ADN et N-OG, un analogue du 2-OG ; A) Représentation en bâtons, la distance entre le groupement méthyle et l'atome d'oxygène de la molécule de NOG est indiquée en angström et ne correspond pas à une liaison hydrogène ; B) Représentation en surface.	26
Figure 15. A) Equilibre tautomérique entre les formes amine et imine de la 5fC et de la 5caC. B) Induction d'un appariement « bancal » par le tautomère imine de la 5caC.	29
Figure 16. Quantification par spectrométrie de masse de la 5hmC (A) et de la 5mC (B) dans différents tissus de souris ⁴⁶ .	30
Figure 17. Variation des niveaux de 5mC et de 5hmC dans le cerveau de souris. A) Coupe sagittale d'un cerveau de souris B) Niveaux de 5hmC et de 5mC dans les différentes régions du cerveau de souris ⁴⁷ .	31
Figure 18. Evolution des niveaux de 5hmC chez les souris au cours du temps. A) Dans le cervelet B) Dans l'hippocampe ⁵⁵ .	33
Figure 19. Evolution des niveaux de 5hmC et de 5mC lors de la mise en culture de cellules issues de tissus mammaires non tumoraux. p0 correspond au début de la culture, p1 au 7 ^e jour de culture, p2 au 14 ^e jour de culture et ainsi de suite. La dernière ligne (α-SS) correspond au contrôle (anticorps contre l'ADN simple brin) ⁴⁵ .	34
Figure 20. Cinétique de la formation de 5hmC (rouge) à partir de 5mC nouvellement formée (bleue) dans des cellules cancéreuses (HCT116 et MCF7) ou des cellules souches embryonnaires de souris (mES). L'ordonnée représente le pourcentage de 5mC ou de 5hmC marquée isotopiquement par rapport à la quantité totale de 5mC ou de 5hmC, respectivement.	35
Figure 21. Facilitation de la déméthylation passive par l'hydroxylation des CpG méthylés. Tiré de ⁷⁶ .	38
Figure 22. Déméthylation de l'ADN par oxydation des 5mC par les protéines TET suivie d'une excision des 5fC ou 5caC par l'enzyme TDG qui va former un site abasique. Celui-ci est ensuite pris en charge par le système BER de réparation de l'ADN. Adapté de ⁷⁷ .	39
Figure 23. Désamination de la 5hmC en 5hmU et excision par une glycosylase spécifique, SMUG1, suivie de la réparation du site abasique obtenu par le système BER.	40

Figure 24. Réactivité en milieu acide et en présence de thiols de A) la 5fC B) la 5caC.....	42
Figure 25. Passage de la thymine à l'uracile, catalysé par la T7H et l'isoorotate décarboxylase.	43
Figure 26. Activités atypiques des DNMT humaines et bactériennes sur la 5mC, la 5hmC et la 5caC observées <i>in vitro</i>	44
Figure 27. Représentation schématique des différentes protéines de la famille MBD et de leurs domaines. D'après ⁹⁷ ; MBD, methyl-CpG binding domain ; TRD transcriptional repression domain; CXXC, domaine CXXC; CC, domaine coiled-coil; G/R, domaine riche en arginines.	48
Figure 28. Aperçu des différentes protéines identifiées lors des expériences de pull-down selon leur affinité pour la 5mC, la 5hmC ou la 5fC.	52
Figure 29. Evolution des niveaux de 5hmC et de 5mC à différentes étapes du développement embryonnaire. A) Dans tout l'embryon ; B) dans les cellules germinales primordiales ²²	54
Figure 30. Evolution du nombre de publications dans le domaine de la 5- hydroxyméthylcytosine depuis les années 1970. (Recherche du mot « 5- hydroxyméthylcytosine » dans la catégorie « Topic », base de donnée « Web of Science Core collections »).	60
Figure 31. Différents axes du projet de thèse.	63
Figure 32. Masse des couples ions parents/fragments des différents nucléosides étudiés par LC-MS/MS MRM.	69
Figure 33. Chromatogrammes UV ou ESI-MS/MS d'un échantillon digéré d'ADN génomique issu de cellules HeLa.	70
Figure 34. Structure d'un inhibiteur irréversible des DNMTs, la 5-azadésoxycytidine.	72
Figure 35. Variation de la méthylation dans différentes lignées cellulaires lors d'un traitement par un inhibiteur de DNMTs. Les graphes A et B représentent les données issues de l'analyse par cytométrie de flux, pour les lignées WM-266 et KG1, respectivement. Les graphes C et D représentent les données issues de l'analyse par spectrométrie de masse.	72
Figure 36. Différents isotopologues utilisées dans la quantification de la 5hmC et la 5mC par la spectrométrie de masse.	74
Figure 37. Spectres de masse et paire ions parents/fragments caractéristiques de A) la désoxycytidine ; B) la 5-méthyl-désoxycytidine ; C) la 5-hydroxyméthyl-désoxycytidine.	75
Figure 38. Chromatogramme LC-ESI-MS/MS d'un mélange des trois nucléosides dC, 5mC et 5hmC.	76
Figure 39. Gamme de concentration des nucléosides.	77
Figure 40. Détection de 5hmC dans un des échantillons de cellules tumorales. Chromatogramme LC-ESI-MS/MS de la A) désoxycytidine B) 5-méthyl-désoxycytidine C) 5- hydroxyméthyl-désoxycytidine.	78
Figure 41. A. Réaction du bisulfite sur la 5hmC conduisant au dérivé CMS ; B. Comparaison de l'efficacité des anticorps anti-5hmC et anti-CMS en immunofluorescence en fonction de la densité de 5hmC par fragment d'ADN ¹²⁴	84
Figure 42. Enrichissement par la méthode hMe-Seal ; A) Transfert d'un motif glucose fonctionnalisable sur la 5hmC par la BGT ; B) Transfert du glucose modifié puis fonctionnalisation avec une biotine.	85
Figure 43. Enrichissement de séquences contenant la 5fC par la méthode fC-Seal ¹⁵⁰	86
Figure 44. Enrichissement de séquences contenant la 5hmC avec la protéine JBP1 ¹⁵²	86
Figure 45. Principe de la méthode GLIB (glucosylation, periodate oxydation and biotinylation) ¹⁴⁸	87
Figure 46. Fonctionnalisation de la 5hmC par un traitement au bisulfite en présence de cystéine-biotine ¹⁵³	88

Figure 47. Enrichissement de séquences contenant la 5fC <i>via</i> la réaction entre un aldéhyde et une hydroxylamine ¹⁵⁴	89
Figure 48. Principe de la cartographie des 5hmC à l'aide de l'enzyme de restriction MspI. .	91
Figure 49. Les produits de réaction des nucléobases C, 5mC, 5hmC, 5caC et 5fC avec le bisulfite de sodium et l'interprétation lors du séquençage (d'après ¹⁶⁰).	93
Figure 50. Adaptation du séquençage au bisulfite à la détection des 5hmC et des 5fC ; A) Principe de la méthode de séquençage OxBS ; B) Principe de la méthode de séquençage RedBS ; C) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.	95
Figure 51. Adaptation du séquençage au bisulfite à la détection des 5hmC ; A) Principe de la méthode de séquençage TAB-seq ¹⁶³ ; B) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.	96
Figure 52. Adaptation du séquençage au bisulfite à la détection des 5fC ; A) Réaction de la 5fC avec l'éthyle-hydroxylamine lors de la méthode fCAB-seq ; B) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.	97
Figure 53. Adaptation du séquençage au bisulfite à la détection des 5caC ; A) Couplage peptidique de la 5caC avec une amine, qui va permettre de la protéger lors du traitement au bisulfite ; B) Récapitulatif des lectures des différentes cytosines modifiées à l'issue du séquençage en fonction de la méthode utilisée.	97
Figure 54. Fonctionnalisation de la 5caC par une biotine par la méthode CAB.	98
Figure 55. Discrimination entre C, 5mC et 5hmC sur un simple brin d'ADN immobilisé dans un pore d' α -hémolysine ¹⁶⁷ . Un brin d'ADN (bleu) est immobilisé dans un pore d' α -hémolysine (grise) à l'aide d'un complexe entre une biotine (jaune) présente à l'extrémité 3' du brin d'ADN et une streptavidine (verte), qui va servir de « bouchon ». Trois sites de reconnaissance (R1, R2 et R3) peuvent permettre de distinguer entre les différentes bases non modifiées. L'introduction de cytosines modifiées dans la séquence d'ADN au site R1, mais pas au site R2, a permis d'observer des variations dans le courant traversant le pore suffisantes pour discriminer en C, 5mC et 5hmC.	99
Figure 56. Séquençage d'un brin d'ADN passant à travers un nanopore ¹⁷⁰ . Un ADN simple brin (ssDNA) est entraîné à travers le pore de MspA (du compartiment trans au compartiment cis) par l'action de l'ADN polymérase ϕ 29. Lorsque les bases passent au niveau du site de reconnaissance de MspA, à proximité du compartiment trans, on peut discriminer les différentes bases de l'ADN, y compris la 5mC et la 5hmC, en fonction des courants spécifiques associés.	100
Figure 57. Principe de la méthode de séquençage SMRT. A) Le complexe polymérase-ADN fixé au sein d'un « zero-mode waveguide ». B) Incorporations successives des nucléotides et lecture des signaux correspondants ¹⁷¹	101
Figure 58. Glucosylation de la 5hmC par la BGT du phage T4.	105
Figure 59. Fonctionnalisation des 5hmC par un sucre comportant une fonction azoture.	106
Figure 60. Méthode enzymatique de quantification de la 5hmC par fluorescence.	107
Figure 61. Séquence du substrat oligonucléotidique contenant une 5hmC.	108
Figure 62. Structure d'un synthon phosphoramidite et des groupes protecteurs utilisés dans la synthèse.	109
Figure 63. Différentes étapes de la synthèse automatique sur phase solide d'ADN.	111
Figure 64. Déprotection totale de l'alcool allylique de la 5hmC.	112
Figure 65. Spectre de masse de l'oligonucléotide contenant la 5hmC.	113
Figure 66. Stratégies générales de synthèse de NDP-sucres.	114
Figure 67. Synthèse du synthon azidoglycosyle phosphate.	116
Figure 68. Synthèse de l'UDP-6-azido-glucose 9	117

Figure 69. Fonctionnalisation du composé 3 par une réaction de chimie « click ».	118
Figure 70. Chromatogrammes HPLC du DBCO-Alexa-Fluor 488 et du milieu réactionnel après 2h à 37°C (Détection à 488 nm).	119
Figure 71. Production et purification de la BGT. A) Principe de la purification de la protéine BGT recombinante. B) Analyse de la purification par gel SDS-PAGE (NI : bactéries transformées où la production de BGT n'a pas été induite ; I : bactéries transformées où la production de BGT a été induite ; 1 : protéine après la première purification ; 2 : protéine après la seconde purification).	120
Figure 72. Représentation de la poche catalytique de la BGT faisant apparaître l'UDP-glucose et la distance entre l'alcool en position 6 du glucose et le résidu Glu 22.	121
Figure 73. Cristal d'un mutant inactif de la BGT (D100A) en présence d'UDP-6-N ₃ -glucose 9 .	122
Figure 74. Fonctionnalisation des 5hmC par pour la quantification par fluorescence.	124
Figure 75. Oxydation sélective de l'alcool allylique de la 5hmC et fonctionnalisation de la 5fC formée.	125
Figure 76. Voies rétrosynthétiques possibles pour obtenir la 5hmdC.	127
Figure 77. Synthèse de la 5-hydroxyméthyl-2'-désoxycytidine à partir de la 2'-désoxyuridine.	128
Figure 78. Oxydation sélective de la 5hmdC en 5fdC	129
Figure 79. Chromatogrammes HPLC de A. la 5hmdC ; des bruts réactionnels de l'oxydation de la 5hmdC par B. RuO ₂ (conditions 3), C. Mn(OAc) ₃ (conditions 5),	131
Figure 80. Fonctionnalisation de la 5fdC par la biotine hydrazine.	132
Figure 81. Fonctionnalisation de la 5fdC A) Profils HPLC de la 5fdC avant réaction ; B) Profils HPLC du brut réactionnel après 12 heures de réaction avec la biotine-hydrazine ; C) Spectre MS/MS de la fraction F3.	133
Figure 82. Oxydation et fonctionnalisation de l'oligonucléotide modèle.	134
Figure 83. Méthode d'enrichissement des 5hmdC avant analyse LC-MS/MS ¹⁸⁸	137
Figure 84. Détection de la 5hmC par une oxydation suivie d'une fonctionnalisation avec une molécule fluorescente.	138
Figure 85. Principes du photomarquage et de l'identification des sites de liaisons. Adapté de ¹⁹³	144
Figure 86. Date d'introduction des différents groupements de photomarquage. D'après ¹⁹⁴	147
Figure 87. Groupements photoactivables les plus couramment utilisés ; (a) benzophénone ; (b) diazirine ; (c) azoture d'aryle.	147
Figure 88. Groupement photoactivables nucléosidiques ; (a) 5-Bromo-2'-désoxyuridine (b) 5-Iodo-2'-désoxyuridine (c) 4-Thiothymidine.	148
Figure 89. Principaux mécanismes de réaction des azotures d'aryles activés	149
Figure 90. Différents analogues de bases de l'ADN contenant un groupement azoture ; a) 5-azidouridine ; b) 8-Azidoadénosine ; c) 2-Azidoadénosine.	149
Figure 91. Principaux mécanismes de réaction de la benzophénone irradiée.	151
Figure 92. Mécanismes réactionnels de l'excitation de la trifluorométhylphényldiazirine.	153
Figure 93. Réaction secondaire ayant lieu après l'insertion du carbène dans une liaison N-H.	154
Figure 94. Voie principale de synthèse des diazirines.	155
Figure 95. Spectres d'absorption de la ⁴⁵ T et de la Thymidine (T). Tiré de ²¹⁵	156
Figure 96. Analogues thiolés des bases de l'ADN. a) 4-Thiouridine ; b) 4-Thiothymidine c) 6-Thio-2'-désoxyguanosine.	156
Figure 97. Dérivés halogénés de l'uridine ; a) 5-Iodo-2'-désoxyuridine b) 5-Iodouridine c) 5-Bromo-2'-désoxyuridine d) 5-Bromouridine.	158

Figure 98. Principe de l'introduction d'une sonde photoactivable à une position spécifique par voie enzymatique. D'après ²³⁴	161
Figure 99. Les deux stratégies principales de fonctionnalisation d'oligonucléotides : A) Marquage pré-synthétique ; B) Marquage post-synthétique. (Les nucléotides sont représentés en bleu ; les fonctions réactives sont représentées sous forme de bâtonnets rouges et les modifications introduites sont représentées sous forme de ronds rouges). D'après ²³⁵	161
Figure 100. Phosphoramidites disponibles commercialement (Glen Research) ; a) phosphoramidite pour l'incorporation de la ⁴⁵ T ; b) phosphoramidite pour l'incorporation de la ⁵¹ dU.....	162
Figure 101. Synthèse du synthon phosphoramidite d'un analogue de thymidine comportant un groupement photoactivable en position 5 du cycle pyrimidique ²³⁸	164
Figure 102. Synthèse et greffage sur colonne d'un analogue de thymidine comportant un groupement photoactivable en position 4' du sucre (LCAA-CPG : Long Chain Aminoalkyl-Controled Pore Glass).	165
Figure 103. Projection des fonctionnalisations dans le grand ou le petit sillon de l'ADN en fonction de la position et de la base fonctionnalisée. D'après ²⁴³	167
Figure 104. Synthèse d'analogues contraints de duplex de l'ADN ; A) Fonctionnalisation post-synthétique de brins d'ADN comportant la 2-Fluoro-2'-désoxyinosine ; B) Structure des sondes et des contrôles utilisées pour étudier l'enzyme M.HaeIII ; Tiré de ²⁴⁴	167
Figure 105. Différents nucléosides convertibles utilisés dans le cadre de la méthode dite du « convertible nucleoside » ; Analogues de G ^{244,246,247} ; Analogues de A ^{246,248} ; Analogues de C ²⁴⁸⁻²⁵⁰ ; Analogues de U ²⁵¹ ; Analogues de T ²⁵⁰	169
Figure 106. Fonctionnalisation post-synthétique de la 2-Fluoro-2'-désoxyinosine et de l'O ⁶ -triazolyl-2'-désoxyuridine par des diazirines aliphatiques.....	170
Figure 107. Conversion post-synthétique de la 4-Triazolo-thymidine en 4-Thiothymidine.	170
Figure 108. Fonctionnalisation post-synthétique de la ⁴⁵ dU en milieu basique par des dérivés halogénés.....	171
Figure 109. Fonctionnalisation post-synthétique de la 5-Iodo-2'-desoxyuridine par une réaction de couplage catalysée au palladium.	172
Figure 110. Fonctionnalisation post-synthétique de la 7-Vinyl-7-deazaguanosine par une réaction de Diels-Alder.	172
Figure 111. Le principe des sondes photoactivables et le rôle des trois modifications introduites.....	174
Figure 112. Choix des modifications introduites dans les sondes photoactivables.....	176
Figure 113. Structure de l'étiquette biotine liée à la colonne.....	177
Figure 114. Phosphoramidite commerciaux utilisés ; a) 2-Fluoro-2'-désoxyinosine ; b) 5-Iodo-2'-désoxyuridine.....	178
Figure 115. Synthèse des amines comportant un groupement photoactivable ; conditions et réactifs ; (a) i) HATU, DIEA, DMF ii) <i>tert</i> -butyl- <i>N</i> -(6-aminohexyl)carbamate 2, DIEA, DMF ; (b) TFA, CH ₂ Cl ₂	179
Figure 116. Rendements de couplage lors de la synthèse automatique (estimés à partir des profils d'absorption UV de déprotection des groupements DMT terminaux). ; A) La barre bleue correspond au couplage de la 5-Iodo-2'-désoxyuridine ; B) La barre orange correspond au couplage de la 2-Fluoro-2'-désoxyinosine. Dans les deux cas, la lettre B correspond à la déprotection de la biotine préchargée sur la colonne, la lettre 5 au couplage du phosphoramidite modifié.....	180
Figure 117. Caractérisation de l'oligonucléotide 16-Bzp-Biot ; A) Spectre de masse obtenu par MALDI-TOF ; B) Spectre HPLC de l'oligonucléotide DMT-off.	182
Figure 118. Structure des modifications introduites dans les oligonucléotides.	183
Figure 119. Liste des sondes utilisées lors de ce chapitre et nomenclature associée.	184

Figure 120. Etude de la stabilité des différents duplex par retard sur gel.	186
Figure 121. Courbe caractéristique permettant de définir la T _m d'un duplex d'ADN (Tiré de http://www.entelechon.com/2008/08/dna-melting-temperature/).	187
Figure 122. Complexation de la Dnmt3a avec Cds-17-Thio ; séparation sur gel natif d'acrylamide (Autoradiogramme).	190
Figure 123. Complexation entre Dnmt3a et Cds-17-Thio ; séparation sur gel natif d'agarose (Autoradiogramme).	191
Figure 124. Photomarquage de DNMT3a par la sonde Cds-17-Thio. Séparation sur gel natif d'acrylamide (Autoradiogramme).	192
Figure 125. Interaction entre hRPA et oligonucléotide simple brin (Autoradiogramme). Le pourcentage de complexation a été calculé en comparant l'intensité de la bande correspondant oligonucléotide-hRPA par rapport à l'intensité totale (oligonucléotide complexé et libre). .	194
Figure 126. Complexation des brins H ou C en présence de la protéine hRPA (Autoradiogramme). Le pourcentage de complexation a été calculé en comparant l'intensité de la bande correspondant au complexe oligonucléotide-hRPA par rapport à l'intensité totale (oligonucléotide complexé et libre).	195
Figure 127. Cinétique d'irradiation des différentes sondes en présence de hRPA (Autoradiogramme, séparation par gel d'acrylamide dénaturant).	197
Figure 128. Evolution du pourcentage de photomarquage au cours du temps en fonction de la nature de la sonde. Exploitation des résultats présentés dans la figure 127. Le pourcentage de photomarquage a été calculé en comparant l'intensité des bandes correspondant aux complexes covalents oligonucléotide-hRPA1 et 2 par rapport à l'intensité totale (oligonucléotide lié covalamment et libre).	198
Figure 129. Spectres d'émission des diodes utilisées dans ce travail, d'après la documentation de Thorlabs ; a) LED300W ; b) LED365W.	201
Figure 130. Utilisation des LEDS pour irradier les sondes 17-Thio, 13-Bzp et 17-Iodo en présence de hRPA (- : pas d'irradiation ; 365, 300 et L représentent respectivement les échantillons irradiés avec la LED à 365 nm, la LED à 300 nm et la lampe UV à 365 nm). La séparation a été effectuée sur un gel d'acrylamide dénaturant (Autoradiogramme).	202
Figure 131. Irradiation d'extraits bactériens surexprimant ou non hRPA (Autoradiogramme). La séparation a été effectuée sur un gel d'acrylamide dénaturant.	204
Figure 132. Irradiation de la sonde 16-Dzr-Biot en présence d'extraits bactériens surexprimant hRPA sauvage ou hRPA mutée (Autoradiogramme). La séparation a été effectuée sur un gel d'acrylamide dénaturant.	206
Figure 133. Western blot des lysats bactériens surexprimant la protéine hRPA (WT) ou le mutant hRPA D228Y (MUT). La quantité de protéines totales des différents extraits a été dosée et la même concentration de protéines a été déposée dans chacun des puits.	207
Figure 134. Photomarquage, capture et relargage de hRPA en présence d'une sonde avec (1-Thio-Biot) ou sans (1-Thio) biotine (Autoradiogramme).	210
Figure 135. Photomarquage d'extraits nucléaires de cellules HeLa par les sonde Cds-16-Dzr-Biot et Hds-16-Dzr-Biot et séparation sur gel dénaturant d'acrylamide (autoradiogramme).	212
Figure 136. Photomarquage d'extraits nucléaires de cerveau de rats par les sondes Cds-16-Dzr-Biot (C), Hds-16-Dzr-Biot (H) et 16-Dzr-Biot (SS).	213

Table des tableaux

Tableau 1. Présence de la 5hmC et des protéines TET dans différents organismes.	36
Tableau 2. Récapitulatif des différentes tentatives de transfert du sucre modifié sur le duplex modèle.	123
Tableau 3. Optimisation des conditions d'oxydation de la 5hmdC	130
Tableau 4. Récapitulatif des différents essais de fonctionnalisation et de purification sur l'oligonucléotide modèle.....	135
Tableau 5. Différentes conditions de fonctionnalisation de l'oligonucléotide testées.....	181
Tableau 6. Valeurs des C_{50} obtenues pour les différentes sondes. C_{50} est la concentration d'ADN complémentaire à laquelle 50% du duplex est formé dans les expériences de gel retard.....	186
Tableau 7. Températures de fusions de différents duplex utilisés dans la suite de l'étude..	188
Tableau 8. Détermination des conditions de relargage des oligonucléotides.	209
Table 9. Mass obtained by MALDI-TOF characterisation of oligodeoxynucleotides.....	242

Résumé :

La 5-hydroxyméthylcytosine (5hmC) est une base modifiée de l'ADN. Chez le mammifère, cette base était considérée jusqu'à récemment comme une espèce mineure produite par des dommages oxydatifs. Cependant, des travaux publiés en 2009 ont profondément changé cette vision. En effet, d'une part, la 5hmC est abondante dans certains types cellulaires, d'autre part la formation de 5hmC est un processus actif, dirigé par des enzymes spécialisées (TET1-3) à partir d'une autre base modifiée de l'ADN, la 5-méthylcytosine. Ces découvertes récentes posent naturellement la question du rôle biologique de la 5hmC. En effet, si la méthylation des cytosines est un phénomène connu dont le rôle primordial dans la régulation de l'expression des gènes est clairement établi, le rôle biologique de la 5hmC est en cours d'exploration et nécessite donc le développement de méthodes dédiées à son étude.

Nous avons pris part au développement de méthodes pour l'étude de l'hydroxyméthylation de l'ADN et des protéines partenaires de cette base en proposant des outils chimiques, biochimiques et analytiques. Ayant montré que la spectrométrie de masse constitué un outil puissant pour la quantification globale de la méthylation de l'ADN, nous avons adapté une méthode utilisant la chromatographie liquide couplée à la spectrométrie de masse à l'étude de l'hydroxyméthylation. Nous avons également participé au développement de méthodes de marquage spécifique des 5hmC selon deux stratégies : (i) une stratégie enzymatique basée sur l'utilisation d'une glucosyltransférase et d'un donneur de glucose (uridine diphosphate glucose) modifié chimiquement permettant l'introduction d'une sonde par une réaction bioorthogonale ; (ii) une stratégie chimique basée sur l'oxydation sélective de l'alcool allylique de la 5hmC. Enfin, nous avons conçu, synthétisé et étudié les propriétés de différentes sondes oligonucléotidiques photoactivables destinées à l'identification de protéines interagissant spécifiquement avec la 5hmC.

Mots clés : Epigénétique, 5-Hydroxyméthylcytosine, Spectrométrie de masse, Photomarquage d'affinité

Chemical tools to study 5hmC, a modified DNA base

5-hydroxymethylcytosine (5hmC) is a modified DNA base, who was until recently believed be a minor modification, resulting of oxidative damage. However, results published in 2009 have challenged this understanding. Indeed, 5hmC is abundant in some cell types and its formation is an active process, mediated by specific enzymes (TET1-3) using another modified DNA base, 5-methylcytosine (5mC). These recent discoveries raise the question of the biological role of 5hmC. Indeed, if the role of 5mC in gene expression regulation is established, the biological role of 5hmC is still in study and require the development of specific methods.

We participated in the development of methods to study 5hmC and its partner proteins using chemical, biochemical and analytical tools. We showed that mass spectrometry is a powerful tool to quantify global methylation levels and adapted a method, using liquid chromatography coupled to tandem mass spectrometry to study hydroxymethylation. We have also taken part in the development of two 5hmC specific labeling methods: (i) an enzymatic strategy, using a glucosyltransferase and a chemically modified glucose donor, allowing the introduction of a probe by a bio-orthogonal reaction; (ii) a chemical strategy based on the selective oxidation of the 5hmC allylic alcohol. Finally, we have designed, synthesized and assessed the properties of photoactivable oligonucleotidic probes in order to identify proteins interacting with 5hmC.

Keywords : Epigenetics, 5-Hydroxymethylcytosine, Mass spectrometry, Photoaffinity labeling